BB-1067-B

TITLE

PLANT METHIONINE SYNTHASE GENE AND METHODS FOR INCREASING THE METHIONINE CONTENT OF THE SEEDS OF PLANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

This is a continuation-in-part of Application No. 08/703,829, filed on August 27, 1996.

TECHNICAL FIELD

This invention relates to a nucleic acid fragment encoding a plant methionine synthase or methionine synthase. The invention also includes chimeric genes, a first encoding a plant methionine synthase (MS) gene, a second encoding a plant cystathionine γ-synthase (CS) gene, a third encoding feedback-insensitive aspartokinase (AK) or bifunctional feedback-insensitive aspartokinase-homoserine dehydrogenase (AK-HDH), which is operably linked to a plant chloroplast transit sequence, and a fourth encoding a methionine-rich protein, all operably linked to plant seed-specific regulatory sequences. Methods for their use to produce increased levels of methionine in the seeds of transformed plants are provided.

BACKGROUND OF THE INVENTION

Human food and animal feed derived from many grains are deficient in the sulfur amino acids, methionine and cysteine, which are required in an animal diet. In corn, the sulfur amino acids are the third most limiting amino acids, after lysine and tryptophan, for the dietary requirements of many animals. The use of soybean meal, which is rich in lysine and tryptophan, to supplement corn in animal feed is limited by the low sulfur amino acid content of the legume. Thus, an increase in the sulfur amino acid content of either corn or soybean would improve the nutritional quality of the mixtures and reduce the need for further supplementation through addition of more expensive methionine.

Efforts to improve the sulfur amino acid content of crops through plant breeding have met with limited success on the laboratory scale and no success on the commercial scale. A mutant corn line which had an elevated whole-kernel methionine concentration was isolated from corn cells grown in culture by selecting for growth in the presence of inhibitory concentrations of lysine plus threonine [Phillips et al., *Cereal Chem.*, (1985), *62*, 213-218]. However, agronomically-acceptable cultivars have not yet been derived from this line and commercialized. Soybean cell lines with increased intracellular concentrations of methionine were isolated by selection for growth in the presence of ethionine [Madison and Thompson, *Plant Cell Reports*, (1988), *7*, 472-476], but plants were not regenerated from these lines.

The amino acid content of seeds is determined primarily by the storage proteins which are synthesized during seed development and which serve as a major nutrient reserve

following germination. The quantity of protein in seeds varies from about 10% of the dry weight in cereals to 20-40% of the dry weight of legumes. In many seeds the storage proteins account for 50% or more of the total protein. Because of their abundance, plant seed storage proteins were among the first proteins to be isolated. Only recently, however, have the amino acid sequences of some of these proteins been determined with the use of molecular genetic techniques. These techniques have also provided information about the genetic signals that control the seed-specific expression and the intracellular targeting of these proteins.

One genetic engineering approach to increase the sulfur amino acid content of seeds is to isolate genes coding for proteins that are rich in the sulfur-containing amino acids methionine and cysteine, to link the genes to strong seed-specific regulatory sequences, to transform the chimeric gene into crops plants and to identify transformants wherein the gene is sufficiently-highly expressed to cause an increase in total sulfur amino acid content. However, increasing the sulfur amino acid content of seeds by expression of sulfur-rich proteins may be limited by the ability of the plant to synthesize methionine, by the synthesis and stability of the methionine-rich protein, and by effects of over-accumulation of the methionine-rich protein on the viability of the transgenic seeds.

An alternative approach would be to increase the production and accumulation of the free amino acid, methionine, via genetic engineering technology. However, little guidance is available on the control of the biosynthesis and accumulation of methionine in plants, particularly in the seeds of plants.

Methionine, along with threonine, lysine and isoleucine, are amino acids derived from aspartate. The first step in the pathway is the phosphorylation of aspartate by the enzyme aspartokinase (AK), and this enzyme has been found to be an important target for regulation of the pathway in many organisms. The aspartate family pathway is also believed to be regulated at the branch-point reactions. For methionine the reduction of aspartyl β-semialdehyde by homoserine dehydrogenase (HDH) may be an important point of control. The first committed step to methionine, the production of cystathionine from *O*-phosphohomoserine and cysteine by cystathionine γ-synthase (CS), appears to be an important point of control of flux through the methionine pathway [Giovanelli et al., *Plant Physiol.*, (1984), *77*, 450-455]. The final step in methionine biosynthesis is catalyzed by the enzyme 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase, also known as methionine synthase (MS). Nucleic acid fragments encoding full-length vitamin-B12 independent methionine synthases from Madagascar periwinkle (*Catharanthus roseus*) [Eichel et al., *Eur. J. Biochem.* (1995), *230,* 1053-1058] Coleus (*Solenostemon scutellarioides*) [Petersen et al., *Plant Physiol.* (1995), *109*, 338], *Arabidopsis thaliana* [Ravanel et al., *Proc. Natl. Acad. Sci. USA* (1998), *95*, 7805-7812], and *Mesembryanthemum crystallinum* [NCBI General Identification No. 1814403], as well as nuceic acid fragments

encoding a portion of vitamin–B12 independent methionine synthase from a number of plant species such as soybean, rice, and corn have been disclosed previously.

SUMMARY OF THE INVENTION

The present invention provides plant genes encoding MS, specifically tobacco, corn and soybean MS genes, additional MS nucleic acid fragments from wheat, as well as chimeric MS genes for seed-specific over-expression of the plant enzyme. Combinations of these genes with other chimeric genes encoding AK, HDH, CS and methionine-rich seed storage protein provide methods to increase the level of methionine in seeds.

More specifically, the present invention concerns an isolated nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code.

In a second embodiment, this invention concerns an isolated nucleic acid fragment comprising:

(a) a first nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code, and

(b) a second nucleic acid fragment encoding a plant cystathionine γ-synthase or a functionally equivalent subfragment thereof.

In a third embodiment, this invention concerns chimeric genes comprising the isolated nucleic acid fragments discussed above operably linked to regulatory sequences.

In a fourth embodiment, this invention concerns plants and transformed hosts comprising such chimeric genes in their genome and seeds obtained from such plants,

In a fifth embodiment, this invention concerns a polypeptide comprising all or a substantial portion of the amino acid sequence set forth in SEQ ID NOS:2, 4, 6, 8 and 10.

In a sixth embodiment, this invention concerns a method for increasing methionine content of the seeds of plants comprising:

(a) transforming plant cells with the chimeric genes discussed above or the nucleic acid fragment discussed above;

(b) growing fertile mature plants from the untransformed plant cells obtained from step (a) under conditions suitable to obtain seeds; and

selecting progeny seed of step (b) for those seeds containing increased levels of methionine compared to untransformed seeds.

In a seventh embodiment, this invention concerns a method for producing plant methionine synthase comprising the following steps:

(a) transforming microbial host cells with a chimeric gene wherein a nucleic acid fragment encoding a plant methionine synthase is operably linked to regulatory sequences capable of expression in microbial cells; then

(b) growing the transformed microbial cells obtained from step (a) under conditions that result in expression of the methionine synthase protein.

In an eighth embodiment,this invention concerns a method for evaluating at least one compound for its ability to inhibit the activity of a plant methionine synthase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a plant methionine synthase, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of plant methionine synthase in the transformed host cell; (c) optionally purifying the plant methionine synthase expressed by the transformed host cell; (d) treating the plant methionine synthase with a compound to be tested; and (e) comparing the activity of the plant methionine synthase that has been treated with a test compound to the activity of an untreated plant methionine synthase, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and the sequence descriptions which form a part of this application.

Figure 1 shows a comparison of the amino acid sequences of *E. coli*, yeast, tobacco, *Catharanthus roseus*, corn and soybean MS proteins.

Figure 2 depicts the amino acid sequence alignment between the methionine synthase encoded by the corn clone p0026.ccras26rb (SEQ ID NO:2), contig assembled from soybean clones s2.17c08, sdc2c.pk001.g7, sdp2c.pk001.e21, sdp2c.pk001.n20, sdp2c.pk012.l21, sdp2c.pk013.d12, sdp2c.pk042.g18, sdp3c.pk001.j3, sdp3c.pk006.n23, sdp3c.pk020.i10, ses4d.pk0010.f10, sfl1.pk129.j22, srm.pk0037.h2 and ssm.pk0070.h6 (SEQ ID NO:4), tobacco clone np.2d06.sk20 (SEQ ID NO:6), wheat clone wlm96.pk0018.c10 (SEQ ID NO:8), wheat clone wl1n.pk0038.e8 (SEQ ID NO:10) and a methionine synthase gene from *Catharanthus roseus* (NCBI General Identifier No. 1362086, SEQ ID NO:11). Amino acids which are conserved among all sequences with an amino acid at that position are indicated with an asterisk (*). Dashes are used by the program to maximize alignment of the sequences.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as

used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1

Methionine Synthase

Protein	Clone Designation	SEQ ID NO: (Nucleotide)	SEQ ID NO: (Amino Acid)
Methionine synthase (corn)	p0026.ccras26rb	1	2
Methionine synthase (soybean)	Contig of s2.17c08 sdc2c.pk001.g7 sdp2c.pk001.e21 sdp2c.pk001.n20 sdp2c.pk012.l21 sdp2c.pk013.d12 sdp2c.pk042.g18 sdp3c.pk001.j3 sdp3c.pk006.n23 sdp3c.pk020.i10 ses4d.pk0010.f10 sfl1.pk129.j22 srm.pk0037.h2 ssm.pk0070.h6	3	4
Methionine synthase (tobacco)	np.2d06.sk20	5	6
Methionine synthase (wheat)	wlm96.pk0018.c10	7	8
Methionine synthase (wheat)	wl1n.pk0038.e8	9	10

SEQ ID NO:11 is the amino acid sequence of a *Catharanthus roseus* methionine synthase NCBI General Identifier No. 1362086.

SEQ ID NOS:12 and 13 set forth the sequences of oligonucleotides that were used in Example 8 to create a BspH I site at the translation start codon of the tobacco methionine synthase gene.

SEQ ID NOS:14 and 15 set forth the sequences of oligonucleotides that were used in Example 8 to create a Kpn I site following the translation stop codon of the tobacco methionine synthase gene.

SEQ ID NO:16 shows the nucleotide sequence of a corn CS cDNA described in Example 3.

SEQ ID NO:17 shows the deduced amino acid sequence of a corn CS protein derived from the nucleotide sequence of SEQ ID NO:16.

SEQ ID NO:18 shows the nucleotide sequence of a 3639 bp Xba I corn genomic DNA fragment encoding two-thirds of the corn CS protein and including 806 bp upstream from the protein coding region as described in Example 3.

SEQ ID NO:19 shows the complete amino acid sequence of the corn CS protein deduced from the corn cDNA genomic DNA fragment of SEQ ID NO:7 and the corn genomic DNA fragment of SEQ ID NO:8.

SEQ ID NOS:20 and 21 show oligonucleotides used to add a translation initiation codon to the corn CS gene.

SEQ ID NO:22 shows the nucleotide sequence of the coding region of the wild type *E. coli* lysC gene, which encodes AKIII, described in Example 5.

SEQ ID NO:23 shows the amino acid sequence of AKIII derived from the nucleotide sequence of SEQ ID NO:22.

SEQ ID NOS:24 and 25 were used in Example 5 to create an Nco I site at the translation start codon of the *E. coli* lysC gene.

SEQ ID NOS:26 and 27 were used in Example 6 to screen a corn library for a high methionine 10 kD zein gene.

SEQ ID NO:28 shows the nucleotide sequence (2123 bp) of the corn HSZ gene. Nucleotides 753-755 are the putative translation initiation codon and nucleotides 1386-1388 are the putative translation termination codon. Nucleotides 1-752 and 1389-2123 include putative 5' and 3' regulatory sequences, respectively.

SEQ ID NO:29 shows the deduced amino acid sequence of the primary translation product of the corn HSZ gene derived from the nucleotide sequence of SEQ ID NO:28.

SEQ ID NOS:30 and 31 were used in Example 7 to modify the HSZ gene by in vitro mutagenesis.

SEQ ID NO:32 shows a 639 bp DNA fragment including the corn HSZ coding region only, which can be isolated by restriction endonuclease digestion using Nco I (5'-CCATGG) to Xba I (5'-TCTAGA). Two Nco I sites that were present in the native HSZ coding region were eliminated by site-directed mutagenesis, without changing the encoded amino acid sequence.

SEQ ID NO:33 shows the deduced amino acid sequence of the corn HSZ protein derived from the nucleotide sequence of SEQ ID NO:32.

SEQ ID NOS:34 and 35 were used in Example 7 to create a form of the HSZ gene with alternative unique restriction endonuclease sites.

SEQ ID NOS:36 and 37 were used in Example 7 to create a gene to code for the mature form of HSZ.

SEQ ID NO:38 shows a 579 bp DNA fragment including the coding region of the mature corn HSZ protein only, which can be isolated by restriction endonuclease digestion using BspH I (5'-TCATGA) to Xba I (5'-TCTAGA). Two Nco I sites that were present in the native HSZ coding region were eliminated by site-directed mutagenesis. This was accomplished without changing the encoded amino acid sequence.

SEQ ID NO:39 shows the deduced amino acid sequence of the corn HSZ protein derived from the nucleotide sequence of SEQ ID NO:38.

SEQ ID NOS:40-45 were used in Example 8 to create a corn chloroplast transit sequence and link the sequence to the *E. coli* lysC-M4 gene.

SEQ ID NOS:46-49 were used in Example 9 to create a soybean chloroplast transit sequence and link the sequence to the *E. coli* lysC-M4 gene.

SEQ ID NOS:50-51 were used in Example 9 and 10 as PCR primers to prepare a DNA fragment carrying the soybean chloroplast transit sequence.

SEQ ID NO:52 was used in Example 9 to remove the corn chloroplast transit sequence from the corn CS gene.

SEQ ID NOS:53-54 were used in Example 10 as PCR primers to isolate and modify the *E. coli* metL gene.

SEQ ID NO:55 shows the nucleotide sequence of a partial soybean MS cDNA, described in Example 1.

The Sequence Descriptions contain the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IYUB standards described in Nucleic Acids Research 13:3021-3030(1985) and in the Biochemical Journal 219 (No. 2):345-373(1984) which are incorporated by reference herein. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

The teachings below describe nucleic acid fragments, chimeric genes and procedures useful for increasing the accumulation of methionine in the seeds of transformed plants, as compared to levels of methionine in untransformed plants.

In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

The terms "subfragment that is functionally equivalent" and "functionally equivalent subfragment" are used interchangeably herein. These terms refer to a portion or subsequence of an isolated nucleic acid fragment in which the ability to alter gene expression or produce a certain phenotype is retained whether or not the fragment or subfragment encodes an active enzyme. For example, the fragment or subfragment can be used in the design of chimeric genes to produce the desired phenotype in a transformed plant. Chimeric genes can be designed for use in co-suppression or antisense by linking a nucleic acid fragment or subfragment thereof, whether or not it encodes an active enzyme, in the appropriate orientation relative to a plant promoter sequence.

The terms "substantially similar" and "corresponding substantially" as used herein refer to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate gene expression or produce a certain phenotype. These terms also refer to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. It is therefore understood, as those skilled in the art will appreciate, that the invention encompasses more than the specific exemplary sequences.

Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under moderately stringent conditions (for example, 0.5 X SSC, 0.1% SDS, 60°C) with the sequences exemplified herein, or to any portion of the nucleotide sequences reported herein and which are functionally equivalent to the nucleic acid fragment of the invention. Preferred substantially similar nucleic acid sequences encompassed by this invention are those sequences that are 80% identical to the nucleic acid fragments reported herein or which are 80% identical to any portion of the nucleotide sequences reported herein. More preferred are nucleic acid fragments which are 90% identical to the nucleic acid sequences reported herein, or which are 90% identical to any portion of the nucleotide sequences reported herein. Most preferred are nucleic acid fragments which are 95 % identical to the nucleic acid sequences reported herein, or which are 95% identical to any portion of the nucleotide sequences reported herein. Sequence alignments and percent similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences are performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are GAP PENALTY=10, GAP LENGTH PENALTY=10, KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4.

A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Altschul, S. F., et al., (1993*) J. Mol. Biol. 215*:403-410) and Gapped Blast (Altschul, S. F. et al., (1997) *Nucleic Acids Res. 25*:3389-3402); see also www.ncbi.nlm.nih.gov/BLAST/). Thus, a substantial portion of an

amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule

would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

As was mentioned above, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent similarity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least 87.6% identical to the amino acid sequence set forth in SEQ ID NO:2, or at least 87.8% identical to the amino acid sequence set forth in SEQ ID NO:4, or at least 92.5% identical to the amino acid sequence set forth in SEQ ID NO:6, or at least 86.3% identical to the amino acid sequence set forth in SEQ ID NO:8, or at least 80% identical to the amino acid sequence set forth in SEQ ID NO:10. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic

acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure. "Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Regulatory

sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include, but are not limited to, promoters, translation leader sequences, introns, and polyadenylation recognition sequences. "Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity.

An "intron" is an intervening sequence in a gene that does not encode a portion of the protein sequence. Thus, such sequences are transcribed into RNA but are then excised and are not translated. The term is also used for the excised RNA sequences. An "exon" is a portion of the sequence of a gene that is transcribed and is found in the mature messenger RNA derived from the gene, but is not necessarily a part of the sequence that encodes the final gene product.

The "translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Molecular Biotechnology 3*:225).

The "3' non-coding sequences" refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid

tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell 1*:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a DNA that is complementary to and synthesized from a mRNA template using the enzyme reverse transcriptase. The cDNA can be single-stranded or converted into the double-stranded form using the klenow fragment of DNA polymerase I. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein within a cell or *in vitro*. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes. The terms "complement" and "reverse complement" are used interchangeably herein with respect to mRNA transcripts, and are meant to define the antisense RNA of the message.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Altered expression" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ significantly from that activity in comparable tissue (organ and of developmental type) from wild-type organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys. 100*:1627-1632).

"End-product inhibition" or "feedback inhibition" refers to a biological regulatory mechanism wherein the catalytic activity of an enzyme in a biosynthetic pathway is reversibly reduced by binding to one or more of the end-products of the pathway when the concentration of the end-product(s) reaches a sufficiently high level, thus slowing the biosynthetic process and preventing over-accumulation of the end-product.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. The preferred method of cell transformation of rice, corn and other monocots is the use of particle-accelerated or "gene gun" transformation technology (Klein et al., (1987) *Nature (London) 327*:70-73; U.S. Patent No. 4,945,050), or an Agrobacterium-mediated method using an appropriate Ti plasmid containing the transgene (Ishida Y. et al., 1996, Nature Biotech. 14:745-750).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Sambrook").

"PCR" or "Polymerase Chain Reaction" is a technique for the synthesis of large quantities of specific DNA segments, consists of a series of repetitive cycles (Perkin Elmer

Cetus Instruments, Norwalk, CT). Typically, the double stranded DNA is heat denatured, the two primers complementary to the 3' boundaries of the target segment are annealed at low temperature and then extended at an intermediate temperature. One set of these three consecutive steps is referred to as a cycle.

An "expression construct"as used herein comprises any of the isolated nucleic acid fragments of the invention used either alone or in combination with each other as discussed herein and further may be used in conjunction with a vector or a subfragment thereof. If a vector is used then the choice of vector is dependent upon the method that will be used to transform host plants as is well known to those skilled in the art. For example, a plasmid vector can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells comprising any of the isolated nucleic acid fragments of the invention. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J. 4*:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis. The terms "expression construct" and "recombinant expression construct" are used interchangeably herein.

The invention concerns an isolated nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code.

Isolation of a plant MS gene

In order to increase the accumulation of free methionine in the seeds of plants via genetic engineering, a gene encoding 5-methyltetrahydropteroyl-triglutamate-homocysteine methyltransferase, also known as methionine synthase (MS), was isolated from several crop plants. MS catalyzes the final reaction in the biosynthesis of methionine.

It is shown that plant MS genes can be isolated and identified by comparison of random plant cDNA sequences to the GenBank database using the BLAST algorithms well known to those skilled in the art. The use of this approach to isolate tobacco, soybean and corn MS cDNA genes is presented in detail in Example 1. The nucleotide sequence of a corn MS cDNA is provided in SEQ ID NO:1, the nucleotide sequence of a soybean MS cDNA assembled from a contig is provided in SEQ ID NO:3, the nucleotide sequence of a tobacco MS cDNA is provided in SEQ ID NO:5, and the partial nucleotide sequence of wheat MS cDNAs is provided in SEQ ID NOS:7 and 9. MS genes from other plants can now be

identified by comparison of random cDNA sequences to the plant MS sequences provided herein. Alternatively, other plant MS genes, either as cDNAs or genomic DNAs, could be isolated directly by using either the tobacco, soybean, corn or wheat MS nucleic acid fragment as a DNA hybridization probe to screen libraries from any desired plant employing methodology well known to those skilled in the art.

Nucleic acid fragments carrying plant MS genes can be used to create chimeric genes which are useful for over-expressing MS in plant cells and in heterologous host cells. When over-expressed in plant cells, either alone or in combination with other proteins described below, MS is useful for increasing the biosynthesis and accumulation of methionine in those cells. It is particularly useful to use the MS gene to increase the methionine content in the cells of the seeds of plants.

It may also be desirable to reduce or eliminate expression of the MS gene in plants for some applications. In order to accomplish this, a chimeric gene designed for cosuppression of MS can be constructed by linking the MS gene or gene fragment to a plant promoter sequences. (See U.S. Patent No. 5,231,020 for methodology to block plant gene expression via cosuppression.) Alternatively, a chimeric gene designed to express antisense RNA for all or part of the MS gene can be constructed by linking the MS gene or gene fragment in reverse orientation to a plant promoter sequences. (See U.S. Patent No. 5,107,065 for methodology to block plant gene expression via antisense RNA.) Either the cosuppression or antisense chimeric gene could be introduced into plants via transformation. Transformants wherein expression of the endogenous MS gene is reduced or eliminated are then selected.

The plant MS protein produced in heterologous host cells, particularly in the cells of microbial hosts, can be used to prepare antibodies to the protein by methods well-known to those skilled in the art. The antibodies are useful for detecting plant MS protein in situ in cells or in vitro in cell extracts. Preferred heterologous host cells for production of plant MS protein are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct chimeric genes for production of plant MS. These chimeric genes could then be introduced into appropriate microorganisms via transformation to provide high level expression of plant MS. An example of a vector for high level expression of plant MS in a bacterial host is provided (Example 2).

In another aspect, this invention concerns a polypeptide comprising all or a substantial portion of the amino acid sequence set forth in SEQ ID NOS:2, 4, 6, 8 and 10.

Additionally, the plant methionine synthase protein can be used as a target to design and/or identify inhibitors of the enzyme that may be useful as herbicides. This is desirable because methionine synthase catalyzes a necessary step in the essential methionine

biosynthetic pathway. Since methionine is metabolized to *S*-adenosyl-methionine, which is used in many important cellular processes, inhibition of methionine biosynthesis results in pleiotropic effects, which potentiate herbicidal activity. Accordingly, inhibition of methionine synthase activity could lead to inhibition of plant growth. Plant methionine synthase differs sufficiently from animal methionine synthase in amino acid sequence and action mechanism (Eichel et al. (1995) *Eur J Biochem 230*:1053-1058; Yamada et al. (1998) *Biosci Biotechnol Biochem 62*:2155-2160) that some inhibitors of plant methionine synthase are likely to be plant-specific. Thus, the instant polypeptides could be appropriate for new herbicide discovery and design.

In still another embodiment, this invention concerns a method for evaluating at least one compound for its ability to inhibit the activity of a plant methionine synthase, the method comprising the steps of:

(a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a plant methionine synthase, operably linked to suitable regulatory sequences;

(b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the plant methionine synthase encoded by the operably linked nucleic acid fragment in the transformed host cell;

(c) optionally purifying the plant methionine synthase expressed by the transformed host cell;

(d) treating the plant methionine synthase with a compound to be tested; and

(e) comparing the activity of the plant methionine synthase that has been treated with a test compound to the activity of an untreated plant methionine synthase,

thereby selecting compounds with potential for inhibitory activity.

Another aspect of the invention concerns an isolated nucleic acid fragment comprising:

(a) a first nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code, and

(b) a second nucleic acid fragment encoding a plant cystathionine γ-synthase or a functionally equivalent subfragment thereof.

Isolation of a plant CS gene

Cystathionine γ-synthase (CS) catalyzes the first reaction wherein cellular metabolites are committed to the synthesis of methionine and has been implicated to play a key role in

the regulation of methionine biosynthesis. Regulation is not achieved through feedback inhibition of CS by any of the pathway end-products [Thompson et al. (1982) *Plant Physiol.* *69*:1077-1083], however. Thus, over-expression of CS is expected to increase flux through the methionine branch of the biosynthetic pathway, even when high levels of methionine are accumulated.

In order to increase the accumulation of free methionine in the seeds of plants it may desirable to increase the expression of cystathionine γ-synthase (CS) in concert with MS. Therefore a gene encoding plant cystathionine γ-synthase (CS) is provided. Also provided herein is a unique nucleic acid fragment containing a plant MS gene linked to a plant CS gene.

A plant CS gene was isolated by complementation of an *E. coli* host strain bearing a metB mutation. Such a strain requires methionine for growth due to inactivation of the *E. coli* gene that encodes CS. Functional expression of the plant CS gene allowed the strain to grow in the absence of methionine. The use of this approach to isolate a corn CS cDNA gene is presented in detail in Example 3. The nucleotide sequence of a corn CS cDNA is provided in SEQ ID NO:16. CS genes from other plants could be similarly isolated by functional complementation of an *E. coli* metB mutation. Alternatively, other plant CS genes, either as cDNAs or genomic DNAs, could be isolated by using the corn CS gene as a DNA hybridization probe.

This invention also concerns a method for increasing methionine content of the seeds of plants comprising:

(a) transforming plant cells with a chimeric gene comprising an isolated nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code operably linked to a regulatory sequence, or a nucleic acid fragment comprising (1) a chimeric gene comprising an isolated nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code operably linked to a regulatory sequence and (2) a second chimeric gene comprising a nucleic acid fragment encoding a plant cystathionine γ-synthase or a functionally equivalent subfragment thereof or a complement thereof operably linked to a regulatory sequence;

(b) growing fertile mature plants from the untransformed plant cells obtained from step (a) under conditions suitable to obtain seeds; and

selecting progeny seed of step (b) for those seeds containing increased levels of methionine compared to untransformed seeds.

Also of interest are plants comprising in their genome such chimeric genes and seeds obtained from such plants.

In still another aspect this invention conerns a method for producing plant methionine synthase comprising:

(a) transforming host cells with a chimeric gene comprising an isolated nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code operably linked to a regulatory sequence, or a nucleic acid fragment comprising (1) a chimeric gene comprising an isolated nucleic acid fragment comprising a nucleotide sequence selected from the group consisting of (a) a nucleotide sequence corresponding to any of the nucleotide sequences set forth in SEQ ID NOS:1, 3, 5, 7 or 9 or the complement thereof, or (b) the nucleotide sequence of (a) wherein said sequence is degenerate in accordance with the degeneracy of the genetic code operably linked to a regulatory sequence and (2) a second chimeric gene comprising a nucleic acid fragment encoding a plant cystathionine γ-synthase or a functionally equivalent subfragment thereof or a complement thereof operably linked to a regulatory sequence;

(b) growing the transformed microbial cells obtained from step (a) under conditions that result in expression of a plant methionine synthase protein.

As is clear from the discussion above, the host cell can be a plant cell or a microbial cell.

Isolation of AK Genes

Over-expression of feedback-insensitive AK increases flux through the entire pathway of aspartate-derived amino acids even in the presence of high concentrations of the pathway end-products lysine, threonine and methionine. This increased flux provides more substrate for CS and MS and increases the potential for methionine over-accumulation.

Provided herein is a unique nucleic acid fragment containing a plant MS gene linked to a plant CS gene and a gene for AK, which is insensitive to feedback-inhibition by end-products of the biosynthetic pathway. Also provided is a unique nucleic acid fragment containing a plant MS gene linked to a plant CS gene and a gene for AK-HDH, both activities of which are insensitive to feedback-inhibition by end-products of the biosynthetic pathway. Over-expression of feedback-insensitive AK-HDH directs the increased flux through the methionine-threonine branch of the aspartate-derived amino acid pathway, further increasing the potential for methionine biosynthesis.

A number of AK and AK-HDH genes have been isolated and sequenced. These include the thrA gene of *E. coli* (Katinka et al. (1980) *Proc. Natl. Acad. Sci.* USA *77*:5730-5733], the metL gene of *E. coli* (Zakin et al. (1983) J. Biol. Chem. 258:3028-3031], the lysC gene of *E. coli* [Cassan et al. (1986) *J. Biol. Chem. 261*:1052-1057], and the HOM3 gene of *S. cerevisiae* [Rafalski et al. (1988) *J. Biol. Chem. 263*:2146-2151]. The thrA gene of *E. coli* encodes a bifunctional protein, AKI-HDHI. The AK activity of this enzyme is inhibited by threonine. The metL gene of *E. coli* also encodes a bifunctional protein, AKII-HDHII, and the AK activity of this enzyme is insensitive to all pathway end-products. The *E. coli* lysC gene encodes AKIII, which is sensitive to lysine inhibition. The HOM3 gene of yeast encodes an AK which is sensitive to threonine.

As indicated above AK genes are readily available to one skilled in the art for use in the present invention. A preferred class of AK genes encoding feedback-insensitive enzymes are derived from the *E. coli* lysC gene. Procedures useful for the isolation of the wild type *E. coli* lysC gene and lysine-insensitive mutations are presented in detail in Example 5.

The sequences of three mutant lysC genes that encoded lysine-insensitive aspartokinase each differed from the wild type sequence by a single nucleotide, resulting in a single amino acid substitution in the protein. Other mutations could be generated at these target sites in vitro by site-directed mutagenesis, using methods known to those skilled in the art. Such mutations would be expected to result in a lysine-insensitive enzyme. Furthermore, the in vivo method described in Example 5 could be used to easily isolate and characterize as many additional mutant lysC genes encoding lysine-insensitive AKIII as desired.

Another preferred class of AK genes are those encoding bi-functional enzymes, AK-HDH, wherein both catalytic activities are insensitive to end-product inhibition. A preferred AK-HDH enzyme is *E. coli* AKII-HDHII encoded by the metL gene. As indicated above, this gene has been isolated and sequenced previously. Thus, it can be easily obtained for use in the present invention by the same method used to obtain the lysC gene described in Example 5. Alternatively, the gene can be isolated from *E. coli* genomic DNA via PCR using oligonucleotide primers designed based on the published DNA sequence as described in Example 9.

In addition to these genes, several plant genes encoding lysine-insensitive AK are known. In barley, lysine plus threonine-resistant mutants bearing mutations in two unlinked genes that result in two different lysine-insensitive AK isoenzymes have been described [Bright et al., *Nature,* (1982), *299,* 278-279, Rognes et al., *Planta,* (1983), *157,* 32-38, Arruda et al., *Plant Phsiol.*, (1984), *76,* 442-446]. In corn, a lysine plus threonine-resistant cell line had AK activity that was less sensitive to lysine inhibition than its parent line [Hibberd et al., *Planta,* (1980), *148,* 183-187]. A subsequently isolated lysine plus threonine-resistant corn mutant is altered at a different genetic locus and also produces

lysine-insensitive AK [Diedrick et al., *Theor. Appl. Genet.*, (1990), *79*, 209-215, Dotson et al., *Planta,* (1990), *182*, 546-552]. In tobacco there are two AK enzymes in leaves, one lysine-sensitive and one threonine-sensitive. A lysine plus threonine-resistant tobacco mutant that expressed completely lysine-insensitive AK has been described [Frankard et al., *Theor. Appl. Genet.*, (1991), *82*, 273-282]. These plant mutants could serve as sources of genes encoding lysine-insensitive AK and used, based on the teachings herein, to increase the accumulation of methionine in the seeds of transformed plants.

A partial amino acid sequence of AK from carrot has been reported [Wilson et al., *Plant Physiol.*, (1991), *97*,1323-1328]. Using this information a set of degenerate DNA oligonucleotides could be designed, synthesized and used as hybridization probes to permit the isolation of the carrot AK gene. Recently the carrot AK gene has been isolated and its nucleotide sequence has been determined [Matthews et al., *U.S.S.N,* (1991), *07*, 746,705]. This gene was used as a heterologous hybridization probe to isolate the *Arabidopsis thaliana* AK-HDH gene [Ghislain et al., *Plant Mol. Biol,* (1994), *24*, 835-851], and thus can be used as a heterologous hybridization probe to isolate the plant genes encoding lysine-insensitive AK or AK-HDH described above.

Methionine-Rich Storage Protein Genes

It may be useful for certain applications to incorporate the excess free methionine produced via deregulation of the biosynthetic pathway into a storage protein. This can help to prevent metabolism of the excess free methionine into such products as S-adenosyl-methionine, which may be undesirable. The storage protein chosen should contain higher levels of methionine than average proteins. Ideally, these methionine-rich storage proteins should contain at least 15% methionine by weight.

A number of methionine-rich plant seed storage proteins have been identified and their corresponding genes have been isolated. A gene in corn for a 15 kD zein protein containing about 15% methionine by weight [Pedersen et al., *J. Biol. Chem.*, (1986), *261*, 6279-6284], a gene for a 10 kD zein protein containing about 30% methionine by weight [Kirihara et al., *Mol. Gen. Genet.*, (1988), *21*, 477-484; Kirihara et al., *Gene*, (1988), *71*, 359-370] have been isolated. A gene from Brazil nut for a seed 2S albumin containing about 24% methionine by weight has been isolated [Altenbach et al., *Plant Mol. Biol.*, (1987), *8*, 239-250]. From rice a gene coding for a 10 kD seed prolamin containing about 25% methionine by weight has been isolated [Masumura et al., *Plant Mol. Biol.*, (1989), *12*, 123-130]. A preferred gene, which encodes the most methionine-rich natural storage protein known, is an 18 kD zein protein designated high sulfur zein (HSZ) containing about 37% methionine by weight that has recently been isolated [World Patent Publication No. WO 92/14822, see Example 6]. Thus, methionine-rich storage protein genes are readily available to one skilled in the art.

Below is provided a discussion concerning the construction of chimeric genes for high-level seed-specific expression of methionine-rich storage protein genes. In addition, there have been several reports on the expression of methionine-rich seed storage protein genes in transgenic plants. The high-methionine 2S albumin from Brazil nut has been expressed in the seeds of transformed tobacco under the control of the regulatory sequences from a bean phaseolin storage protein gene. The protein was efficiently processed from a 17 kD precursor to the 9 kD and 3 kD subunits of the mature native protein. The accumulation of the methionine-rich protein in the tobacco seeds resulted in an up to 30% increase in the level of methionine in the seeds [Altenbach et al., *Plant Mol. Biol.*, (1989), *13*, 513-522]. This methionine-rich storage protein has also been efficiently expressed in Canola seeds [Altenbach et al., *Plant Mol. Biol.*, (1992), *18*, 235-245.] In another case, high-level seed-specific expression of the 15 kD methionine-rich zein, under the control of the regulatory sequences from a bean phaseolin storage protein gene, was found in transformed tobacco; the signal sequence of the monocot precursor was also correctly processed in these transformed plants [Hoffman et al., *EMBO J.*, (1987), *6*, 3213-3221]. As another example, the 18 kD zein protein containing 37% methionine has been expressed in tobacco and soybean seeds [World Patent Publication No. WO 92/14822].

Construction of Chimeric Genes for Expression of
MS, CS, AK, AK-HDH and methionine-rich storage proteins in the Seeds of Plants

In order to increase biosynthesis of methionine in seeds, suitable regulatory sequences are provided to create chimeric genes for high level seed-specific expression of the MS, CS, AK or AK-HDH and methionine-rich storage proteins. The replacement of the native regulatory sequences accomplishes three things: 1) any methionine-concentration-dependent regulatory sequences are removed, permitting biosynthesis to continue in the presence of high levels of free methionine, 2) any pleiotropic effects that the accumulation of excess free methionine might have on the vegetative growth of plants is prevented because the chimeric gene(s) is not expressed in vegetative tissue of the transformed plants 3) high level expression of the enzyme(s) and storage protein(s) is obtained in the seeds.

The expression of foreign genes in plants is well-established [De Blaere et al., *Meth. Enzymol.*, (1987), *143*, 277-291]. Proper level of expression of MS, CS, AK or AK-HDH and methionine-rich storage protein mRNAs may require the use of different chimeric genes utilizing different promoters. Such chimeric genes can be transferred into host plants either together in a single expression vector or sequentially using more than one vector. Preferred among the higher plants and the seeds derived from them are soybean, rapeseed (*Brassica napus, B. campestris*), sunflower (*Helianthus annus*), cotton (*Gossypium hirsutum*), corn, tobacco (*Nicotiana Tubacum*), alfalfa (*Medicago sativa*), wheat (*Triticum* sp.), barley (*Hordeum vulgare*), oats (*Avena sativa*, L), sorghum (*Sorghum bicolor*), rice (*Oryza sativa*),

and forage grasses. Expression in plants will use regulatory sequences functional in such plants.

The origin of the promoter chosen to drive the expression of the coding sequence is not critical as long as it has sufficient transcriptional activity to accomplish the invention by expressing translatable mRNA for MS, CS, AK or AK-HDH and methionine-rich storage protein genes in the desired host tissue.

Preferred promoters are those that allow expression of the protein specifically in seeds. This may be especially useful, since seeds are the primary source of vegetable amino acids and also since seed-specific expression will avoid any potential deleterious effect in non-seed organs. Examples of seed-specific promoters include, but are not limited to, the promoters of seed storage proteins. The seed storage proteins are strictly regulated, being expressed almost exclusively in seeds in a highly organ-specific and stage-specific manner [Higgins et al., *Ann. Rev. Plant Physiol.*, (1984), *35*, 191-221; Goldberg et al., *Cell,* (1989), *56*, 149-160; Thompson et al., *BioEssays,* (1989), *10*, 108-113]. Moreover, different seed storage proteins may be expressed at different stages of seed development.

There are currently numerous examples for seed-specific expression of seed storage protein genes in transgenic dicotyledonous plants. These include genes from dicotyledonous plants for bean β-phaseolin [Sengupta-Goplalan et al., *Proc. Natl. Acad. Sci. USA*, (1985), *82*, 3320-3324; Hoffman et al., *Plant Mol. Biol.*, (1988), *11*, 717-729], bean lectin [Voelker et al., *EMBO J.*, (1987), *6*, 3571-3577], soybean lectin [Okamuro et al., *Proc. Natl. Acad. Sci. USA*, (1986), *83*, 8240-8244], soybean kunitz trypsin inhibitor [Perez-Grau et al., *Plant Cell*, (1989), *1*, 095-1109], soybean β–conglycinin [Beachy et al., *EMBO J.*, (1985), *4*, 3047-3053; Barker et al., *Proc. Natl. Acad. Sci. USA*, (1988), *85*, 458-462; Chen et al., *EMBO J.*, (1988), *7*,297-302; Chen et al., *Dev. Genet.*, (1989), *10*, 112-122; Naito et al., *Plant Mol. Biol.*, (1988), *11*, 109-123], pea vicilin [Higgins et al., *Plant Mol. Biol.*, (1988), *11*, 683-695], pea convicilin [Newbigin et al., *Planta*, (1990), *180*, 461], pea legumin [Shirsat et al., *Mol. Gen. Genetics*, (1989), *215*, 326]; rapeseed napin [Radke et al., *Theor. Appl. Genet.*, (1988), *75*, 685-694] as well as genes from monocotyledonous plants such as for maize 15 kD zein [Hoffman et al., *EMBO J.*, (1987), *6*, 3213-3221; Schernthaner et al., *EMBO J.*, (1988), *7*, 1249-1253; Williamson et al., *Plant Physiol.*, (1988), *88*, 1002-1007], barley β-hordein [Marris et al., *Plant Mol. Biol.*, (1988), *10*, 359-366] and wheat glutenin [Colot et al., *EMBO J.*, (1987), *6*, 3559-3564]. Moreover, promoters of seed-specific genes, operably linked to heterologous coding sequences in chimeric gene constructs, also maintain their temporal and spatial expression pattern in transgenic plants. Such examples include linking either the Phaseolin or Arabidopsis 2S albumin promoters to the Brazil nut 2S albumin coding sequence and expressing such combinations in tobacco, Arabidopsis, or Brassica napus [Altenbach et al., *Plant Mol. Biol.*, (1989), *13*, 513-522; Altenbach et al., *Plant Mol. Biol.*, (1992), *18*, 235-245; De Clercq et al., *Plant Physiol.*, (1990), *94*, 970-979],

bean lectin and bean β-phaseolin promoters to express luciferase [Riggs et al., *Plant Sci.*, (1989), *63*, 47-57], and wheat glutenin promoters to express chloramphenicol acetyl transferase [Colot et al., *EMBO J.*, (1987), *6*, 3559-3564].

Of particular use in the expression of the nucleic acid fragment of the invention will be the heterologous promoters from several extensively-characterized soybean seed storage protein genes such as those for the Kunitz trypsin inhibitor [Jofuku et al., *Plant Cell*, (1989), *1*, 1079-1093; Perez-Grau et al., *Plant Cell*, (1989), *1*, 1095-1109], glycinin [Nielson et al., *Plant Cell*, (1989), *1*, 313-328], β-conglycinin [Harada et al., *Plant Cell*, (1989), *1*, 415-425]. Promoters of genes for α′- and β-subunits of soybean β-conglycinin storage protein will be particularly useful in expressing the CS, AK and AK-HDH mRNAs in the cotyledons at mid- to late-stages of soybean seed development [Beachy et al., *EMBO J.*, (1985), *4*, 3047-3053; Barker et al., *Proc. Natl. Acad. Sci. USA*, (1988), *85*, 458-462; Chen et al., *EMBO J.*, (1988), *7*, 297-302; Chen et al., *Dev. Genet.*, (1989), *10*, 112-122; Naito et al., *Plant Mol. Biol.*, (1988), *11*, 109-123] in transgenic plants, since: a) there is very little position effect on their expression in transgenic seeds, and b) the two promoters show different temporal regulation: the promoter for the α′-subunit gene is expressed a few days before that for the β-subunit gene.

Also of particular use in the expression of the nucleic acid fragments of the invention will be the promoters from several extensively characterized corn seed storage protein genes such as endosperm-specific promoters from the 10 kD zein [Kirihara et al., *Gene*, (1988), *71*, 359-370], the 27 kD zein [Prat et al., *Gene*, (1987), *52*, 51-49; Gallardo et al., *Plant Sci.*, (1988), *54*, 211-281], and the 19 kD zein [Marks et al., *J. Biol. Chem.*, (1985), *260*, 16451-16459]. The relative transcriptional activities of these promoters in corn have been reported [Kodrzyck et al., *Plant Cell*, (1989), *1*, 105-114] providing a basis for choosing a promoter for use in chimeric gene constructs for corn. For expression in corn embryos, the strong embryo-specific promoter from the GLB1 gene [Kriz, *Biochemical Genetics*, (1989), *27*, 239-251, Wallace et al., *Plant Physiol.*, (1991), *95*, 973-975] can be used.

It is envisioned that the introduction of enhancers or enhancer-like elements into other promoter constructs will also provide increased levels of primary transcription for MS, CS, AK or AK-HDH and methionine-rich storage protein genes to accomplish the invention. These would include viral enhancers such as that found in the 35S promoter [Odell et al., *Plant Mol. Biol.*, (1988), *10*, 263-272], enhancers from the opine genes [Fromm et al., *Plant Cell*, (1989), *1*, 977-984], or enhancers from any other source that result in increased transcription when placed into a promoter operably linked to the nucleic acid fragment of the invention.

Of particular importance is the DNA sequence element isolated from the gene for the α′-subunit of β-conglycinin that can confer 40-fold seed-specific enhancement to a constitutive promoter [Chen et al., *EMBO J*, (1988), *7*, 297-302; Chen et al., *Dev. Genet.*,

(1989), *10*, 112-122]. One skilled in the art can readily isolate this element and insert it within the promoter region of any gene in order to obtain seed-specific enhanced expression with the promoter in transgenic plants. Insertion of such an element in any seed-specific gene that is expressed at different times than the β–conglycinin gene will result in expression in transgenic plants for a longer period during seed development.

Any 3' non-coding region capable of providing a polyadenylation signal and other regulatory sequences that may be required for the proper expression of the CS and AK coding regions can be used to accomplish the invention. This would include the 3' end from any storage protein such as the 3' end of the bean phaseolin gene, the 3' end of the soybean β–conglycinin gene, the 3' end from viral genes such as the 3' end of the 35S or the 19S cauliflower mosaic virus transcripts, the 3' end from the opine synthesis genes, the 3' ends of ribulose 1,5-bisphosphate carboxylase or chlorophyll a/b binding protein, or 3' end sequences from any source such that the sequence employed provides the necessary regulatory information within its nucleic acid sequence to result in the proper expression of the promoter/coding region combination to which it is operably linked. There are numerous examples in the art that teach the usefulness of different 3' non-coding regions [for example, see Ingelbrecht et al., *Plant Cell*, (1989), *1*, 671-680].

DNA sequences coding for intracellular localization sequences may be added to the AK or AK-HDH coding sequence if required for the proper expression of the proteins to accomplish the invention. Plant amino acid biosynthetic enzymes are known to be localized in the chloroplasts and therefore are synthesized with a chloroplast targeting signal. The plant-derived MS, CS and methionine-rich storage protein coding sequences include the native intracellular targeting signals, but bacterial proteins such as *E. coli* AKIII and AKII-HDHII have no such signal. A chloroplast transit sequence could, therefore, be fused to the coding sequence. Preferred chloroplast transit sequences are those of the small subunit of ribulose 1,5-bisphosphate carboxylase, e.g. from soybean [Berry-Lowe et al., *J. Mol. Appl. Genet*, (1982), *1*, 483-498] for use in dicotyledonous plants and from corn [Lebrun et al., *Nucleic Acids Res.*, (1987), *15*, 4360] for use in monocotyledonous plants.

Introduction of Chimeric Genes into Plants

Various methods of introducing a DNA sequence into eukaryotic cells (i.e., of transformation) of higher plants are available to those skilled in the art (see EPO publications 0 295 959 A2 and 0 138 341 A1). Such methods include those based on transformation vectors utilizing the Ti and Ri plasmids of *Agrobacterium* spp. It is particularly preferred to use the binary type of these vectors. Ti-derived vectors transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants, such as soybean, cotton and rape [Pacciotti et al., *Bio/Technology*, (1985), *3*, 241; Byrne et al., *Plant Cell, Tissue and Organ Culture*, (1987), *8*, 3; Sukhapinda et al., *Plant Mol. Biol.*, (1987), *8*,209-216; Lorz et al., *Mol. Gen. Genet.*, (1985), *199*, 178; Potrykus, *Mol. Gen. Genet.*, (1985), *199*, 183].

Other transformation methods are available to those skilled in the art, such as direct uptake of foreign DNA constructs [see EPO publication 0 295 959 A2], techniques of electroporation [see Fromm et al., *Nature (London)*, (1986), *319*, 791] or high-velocity ballistic bombardment with metal particles coated with the nucleic acid constructs [see Kline et al., *Nature (London)*, (1987), *327*, 70, and see U.S. Patent No. 4,945,050]. Once transformed, the cells can be regenerated by those skilled in the art.

Of particular relevance are the methods to transform foreign genes into commercially important crops, such as rapeseed [see De Block et al., *Plant Physiol.*, (1989), *91*, 694-701], sunflower [Everett et al., *Bio/Technology*, (1987), *5*, 1201], soybean [McCabe et al., *Bio/Technology*, (1988), *6*, 923; Hinchee et al., *Bio/Technology*, (1988), *6*, 915; Chee et al., *Plant Physiol.*, (1989), *91*, 1212-1218; Christou et al., *Proc. Natl. Acad. Sci USA*, (1989), *86*, 7500-7504; EPO Publication 0 301 749 A2], and corn [Gordon-Kamm et al., *Plant Cell*, (1990), *2*, 603-618; Fromm et al, *Biotechnology*, (1990), *8*, 833-839].

There are a number of methods that can be used to obtain nucleic acid fragments and plants containing multiple chimeric genes of this invention. Chimeric genes for seed-specific expression of MS, CS, AK or AKHDH and methionine-rich storage proteins can be linked on a single nucleic acid fragment which can be used for transformation. Plants wherein two or more chimeric genes are linked on a nucleic acid fragment integrated into a plant chromosome are selected. In another method two or more of the MS, CS, AK or AKHDH and methionine-rich storage protein chimeric genes, carried on separate DNA fragments, are co-transformed into the target plant and transgenic plants carrying two or more chimeric genes linked on a nucleic acid fragment integrated into a plant chromosome are selected. Alternatively, a plant transformed with an MS chimeric gene can be crossed with a plant transformed with a CS, AK or AKHDH and/or a methionine-rich storage protein chimeric gene, and hybrid plants carrying two or more chimeric genes can be selected. In yet another method a plant transformed with one of the chimeric genes is re-transformed with another chimeric gene or genes.

Expression of Chimeric Genes in Transformed Plants

To analyze for expression of the chimeric MS, CS, AK, AK-HDH and methionine-rich storage protein gene in seeds and for the consequences of expression on the amino acid content in the seeds, a seed meal can be prepared by any suitable method. The seed meal can be partially or completely defatted, via hexane extraction for example, if desired. Protein extracts can be prepared from the meal and analyzed for MS, CS, AK or HDH enzyme activities. Alternatively the presence of any of the proteins can be tested for immunologically by methods well-known to those skilled in the art. To measure free amino acid composition of the seeds, free amino acids can be extracted from the meal by methods known to those skilled in the art [for example, Bieleski et al., *Anal. Biochem.*, (1966), *17*, 278-293]. Amino acid composition can then be determined using any commercially

available amino acid analyzer. To measure total amino acid composition of the seeds, meal containing both protein-bound and free amino acids can be acid-hydrolyzed to release the protein-bound amino acids and the composition can then be determined using any commercially available amino acid analyzer. Seeds expressing the MS, CS, AK, AK-HDH and/or methionine-rich storage proteins and with higher methionine content than the wild type seeds can thus be identified and propagated.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Isolation of Plant MS Genes

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
np	Young Tobacco Green Seedling	np.2d06.sk20
p0026	Corn Regenerating Callus 5 Days After Auxin Removal	p0026.ccras26rb
s2	Soybean Seed, 19 Days After Flowering	s2.17b10 s2.17c08
sdc2c	Soybean Developing Cotyledon (6-7 mm)	sdc2c.pk001.g7
sdp2c	Soybean Developing Pod (6-7 mm)	sdp2c.pk001.e21 sdp2c.pk001.n20 sdp2c.pk012.l21 sdp2c.pk013.d12 sdp2c.pk042.g18
sdp3c	Soybean Developing Pod (8-9 mm)	sdp3c.pk001.j3 sdp3c.pk006.n23 sdp3c.pk020.i10
ses4d	Soybean Embryogenic Suspension 4 Days After Subculture	ses4d.pk0010.f10
sfl1	Soybean Immature Flower	sfl1.pk129.j22
srm	Soybean Root Meristem	srm.pk0037.h2
ssm	Soybean Shoot Meristem	ssm.pk0070.h6
wl1n	Wheat Leaf From 7 Day Old Seedling*	wl1n.pk0038.e8
wlm96	Wheat Seedling 96 Hours After Inoculation With *Erysiphe graminis f. sp tritici*	wlm96.pk0018.c10

*This library was normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP* XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP* XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer

sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science 252*:1651). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer. Complete nucleotide sequence of the cDNAs may be determined using a ABI Model 373A DNA sequencer.

cDNA clones encoding methionine synthase were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nature Genetics 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

A tobacco cDNA library was constructed using RNA derived from young green seedlings. The RNA was sent to Stratagene Cloning Systems (La Jolla, CA) for the custom synthesis of a cDNA library in the Lambda Uni-Zap™ XR vector. Randomly picked individual cDNA inserts were amplified from phage DNA via PCR and the DNA was sequenced using a ABI Model 373A DNA sequencer. The DNA sequences were analyzed for similarity to all publicly available previous DNA sequences in the GeneBank Database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available previous protein sequences in the GeneBank Database using the BLASTX algorithm provided by the NCBI.

The BLASTX search using clone np.2d06.sk20 revealed unmistakable similarity of the protein encoded by the DNA to *E. coli* MS and yeast MS. The amino acid sequence similarity began essentially at the start of both the *E. coli* and yeast proteins indicating that the tobacco cDNA was likely to be a nearly full length cDNA. *E. coli* MS is a protein of 753 amino acids and yeast MS contains 767 amino acids. Thus, the coding region of the tobacco MS would be expected to be 2250-2300 nucleotides long. A plasmid-borne vector carrying the tobacco MS cDNA insert was excised from the lambda phage using the standard lambda-

zap procedure provided by Stratagene and designated pBT771. The ampicillin-resistant plasmid carried the cDNA insert in the vector pBluescript SK(-). Restriction endonuclease digests of the plasmid indicated that the cDNA insert was about 2.6 kb, thus long enough to encode a complete tobacco MS protein.

The complete nucleotide sequence of the full length tobacco MS cDNA clone was determined using a ABI Model 373A DNA sequencer. SEQ ID NO:5 shows the nucleotide sequence of the tobacco MS cDNA and the corresponding amino acid sequence of the tobacco MS protein. The amino acid sequence of tobacco MS shows approximately 44% sequence similarity to either the *E. coli* or yeast proteins.

Similarly, a corn cDNA library was constructed using RNA derived from developing kernels 15 days after pollination. The RNA was sent to Stratagene Cloning Systems (La Jolla, CA) for the custom synthesis of a cDNA library in the Lambda Uni-Zap™ XR vector, randomly picked individual cDNA inserts were amplified from phage DNA via PCR, and the DNA was sequenced using a ABI Model 373A DNA sequencer. The DNA sequences were analyzed as described above.

The BLASTX search using clone m.15.4.c03.sk20 revealed unmistakable similarity of the protein encoded by the DNA to *E. coli* MS, yeast MS and tobacco MS (Figure 1). The amino acid sequence similarity began in the middle of both the *E. coli* and yeast proteins indicating that the corn cDNA was not a full length cDNA. The partial corn MS amino acid sequence shows 45% similarity to *E. coli* MS, 48% similarity to yeast MS and 61% similarity to tobacco MS. Using similar methods, another corn cDNA clone, p0026.ccras26rb, was found to encode a full-length methionine synthase. The sequence of this corn MS cDNA is shown in SEQ ID NO:1.

Soybean cDNA libraries were constructed using RNA derived from developing seeds. The RNA was used to create cDNA libraries in the Lambda™Uni-Zap XR vector; bacterial plasmid cDNA libraries were derived from the phages using the Lambda-Zap procedure. Randomly picked individual cDNA clones were amplified from bacterial DNA via PCR, using primers complimentary to plasmid DNA that flanked the cDNA insert. The DNA was sequenced using a ABI Model 377 DNA sequences, and the DNA sequences were analyzed as described above.

The BLASTX search using clones s2.09c02, s2.11g02, s2.17b10, s2.17c08, s2.17f07 and se2.11f12 each revealed unmistakably similarity of the protein encoded by the DNA to *E. coli* MS, yeast MS, tobacco MS and *Catharanthus roseus* MS. A contiguous sequence was constructed using the 6 cDNA sequences above. The sequence of this soybean MS cDNA is shown in SEQ ID NO:55. The deduced amino acid sequence of the soybean cDNA showed strong similarity to the carboxy half of the *E. coli*, yeast and especially the tobacco and *Catharanthus roseus* MS proteins (Figure 1). A longer contig from various soybean

clones (Table 1) was assembled to encode a full-length methionine synthase, the nucleotide sequence of which is shown in SEQ ID NO:3.

The BLASTX search using the sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to methionine synthase from different plant species. Shown in Table 3 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), or contigs assembled from two or more ESTs ("Contig"):

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous to Methionine Synthase

		BLAST Results		
Clone	Status	Organism	General Identification No.	pLog Score
p0026.ccras26rb	FIS	Catharanthus roseus	1362086	>254.00
Contig of s2.17c08 sdc2c.pk001.g7 sdp2c.pk001.e21 sdp2c.pk001.n20 sdp2c.pk012.l21 sdp2c.pk013.d12 sdp2c.pk042.g18 sdp3c.pk001.j3 sdp3c.pk006.n23 sdp3c.pk020.i10 ses4d.pk0010.f10 sfl1.pk129.j22 srm.pk0037.h2 ssm.pk0070.h6	Contig	Arabidopsis thaliana	2738248	>254.00
np.2d06.sk20	FIS	Catharanthus roseus	1362086	>254.00
wlm96.pk0018.c10	EST	Mesembryanthemum crystallinum	1814403	70.40
wl1n.pk0038.e8	EST	Catharanthus roseus	1362086	29.30

Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8 and 10 and the *Catharanthus roseus* sequence (SEQ ID NO:11). The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8 and 10 and the *Catharanthus roseus* sequence (SEQ ID NO:11).

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Methionine Synthase

SEQ ID NO.	Percent Identity to General Identification No. 1362086; SEQ ID NO:11
2	87.5
4	87.5
6	92.4
8	84.7
10	57.8

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a methionine synthase. These sequences represent the first monocot (corn), soybean and tobacco cDNA sequences encoding full-length methionine synthase as well as the first wheat partial cDNA sequences encoding methionine synthase.

EXAMPLE 2

Construction of Chimeric MS Genes for Expression in *E. coli*

The tobacco MS gene was modified to permit the construction of chimeric genes for expression in *E. coli* and plant seeds. First, a BspH I site (TCATGA) was introduced at the ATG start codon using oligonucleotides CF49 and (SEQ ID NO:12) CF50 (SEQ ID NO:13). The oligonucleotides were annealed and inserted into pBT771 digested with EcoR I and EcoR V. This takes advantage of the unique EcoR I site at the junction of the vector and cDNA and a unique EcoR V site about 20bp from the start codon. The result of this insertion is to remove the cDNA sequences upstream of the ATG start codon and to alter the second codon of the tobacco MS gene from GCA encoding alanine to ACA encoding threonine. Since threonine is the second amino acid in *E. coli* MS, the substitution of threonine for alanine in tobacco MS is not expected to affect the protein function. Insertion of an oligonucleotide with the correct sequence into this region was confirmed by DNA sequencing, yielding plasmid pBT772.

Next, a Kpn I site was added immediately following the translation stop codon. This was accomplished by using PCR employing pBT771 DNA as template and primers CF51

(SEQ ID NO:14) and CF52 (SEQ ID NO:15) to generate a modified 280 base pair fragment that was digested with Rsr II and Kpn I and inserted into similarly digested pBT772. This DNA fragment replacement removes the 3' non-coding sequences present in the tobacco MS cDNA. Insertion of a DNA fragment with the correct sequence was confirmed by DNA sequencing, yielding plasmid pBT773, which contains a unique 2306 bp BspH I to Kpn I fragment carrying the tobacco MS coding region only.

To achieve high level expression of the tobacco MS gene in *E. coli* a modified version bacterial expression vector pET-3a [Rosenberg et al., *Gene*, (1987), *56*, 125-135] was used. This expression vector employs the bacteriophage T7 RNA polymerase/T7 promoter system. First, an oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG creating pBT430. Then pBT430 was further modified to include a Kpn I site downstream of the Nco I site at the translation initiation codon. The tobacco MS gene was cut out of pBT773 as a 2300 bp BspH I-Kpn I fragment and inserted into the above described expression vector digested with Nco I and Kpn I.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 3

Isolation of a Plant CS Gene

In order to clone the corn CS gene, RNA was isolated from developing seeds of corn line H99 19 days after pollination. This RNA was sent to Clontech Laboratories, Inc., (Palo Alto, CA) for the custom synthesis of a cDNA library in the vector Lambda Zap II. The conversion of the Lambda Zap II library into a phagemid library, then into a plasmid library was accomplished following the protocol provided by Clontech. Once converted into a

plasmid library the ampicillin-resistant clones obtained carry the cDNA insert in the vector pBluescript SK(-). Expression of the cDNA is under control of the lacZ promoter on the vector.

Two phagemid libraries were generated using the mixtures of the Lambda Zap II phage and the filamentous helper phage of 100 mL to 1 mL. Two additional libraries were generated using mixtures of 100 μL Lambda Zap II to 10 μL helper phage and 20 μL Lambda Zap II to 10 μL helper phage. The titers of the phagemid preparations were similar regardless of the mixture used and were about 2 x 10^3 ampicillin-resistant-transfectants per mL with *E. coli* strain XL1-Blue as the host.

To identify clones that carried the CS gene, *E. coli* strain BOB105 was constructed by introducing the F' plasmid from *E. coli* strain XL1-blue into strain UB 1005 [Clark (1984) FEMS Microbiol. Lett. 21:189] by conjugation. The genotype of BOB105 is: F'::Tn10 $proA^+B^+$ $lacI^q$ D(lacZ)M15/nalA37 metB1. The strain requires methionine for growth due to a mutation in the metB gene that encodes CS. Functional expression of the plant CS gene should complement the mutation and allow the strain to grow in the absence of methionine.

To select for clones from the corn cDNA library that carried the CS gene, 100 μL of the phagemid library was mixed with 300 μL of an overnight culture of BOB105 grown in L broth and incubated at 37° for 15 min. The cells were collected by centrifugation, resuspended in 400 μL of M9 + vitamin B1 broth and plated on M9 media containing vitamin B1, glucose as a carbon and energy source, 20 μL threonine (to prevent the possibility of threonine starvation due to overexpression of CS), 100 μg/mL ampicillin, 20 μg/mL tetracycline, and 0.16 mM IPTG (isopropylthio-β-galactoside). Fifteen plates were prepared and incubated at 37°C. The amount of phagemid added was expected to yield about 2 x 10^5 ampicillin-resistant transfectants per plate.

Approximately 30 colonies (an average of 2 per plate or 1 per 10^5 transfectants) able to grow in the absence of methionine were obtained. No colonies were observed if the phagemids carrying the corn cDNA library were not added. Twelve clones were picked and colony purified by streaking on the same medium described above. Plasmid DNA was isolated from the 12 clones and retransformed into BOB105. All of the 12 DNAs yielded methionine-independent transformants demonstrating that a plasmid-borne gene was responsible for the phenotype. The sequence of the DNA insert in one of the plasmids, FS1088, is shown in SEQ ID NO:16 and the deduced amino acid sequence of the corn CS protein derived from the nucleotide sequence of SEQ ID NO:16 is shown in SEQ ID NO:17. It is 1639 bp in length and contains a long open reading frame and a poly A tail, indicating that it too represents a corn cDNA. The deduced amino acid sequence of the open reading frame shows 59 percent similarity and 34 percent identity to the published sequence of *E. coli* CS, indicating that it represents a corn homolog to the *E. coli* metB gene.

The open reading frame in plasmid FS1088 continues to the end of the insert DNA, and does not include an ATG initiator codon, indicating that the cloned cDNA is incomplete. The open reading frame of FS1088 is in frame with the initiator codon of the lacZ gene carried on the cloning vector. Thus, complementation of the metB mutation in BOB105 results from expression of a fusion protein including 39 amino acids from β-galactosidase and the vector polylinker attached to the truncated corn CS protein.

In order to clone the entire 5' end of the corn CS gene, the cDNA clone was used as a DNA hybridization probe to screen a genomic corn library. A genomic library of corn in bacteriophage lambda was purchased from Stratagene (La Jolla, California). Data sheets from the supplier indicated that the corn DNA was from etiolated Missouri 17 corn seedlings. The vector was Lambda FIX™ II carrying Xho 1 fragments 9-23 kb in size. A titer of 1.0×10^{10} plaque forming units (pfu)/mL in the amplified stock was indicated by the supplier when purchased. Prior to screening, the library was re-titered and contained 2.0×10^8 pfu/mL.

The protocol used for screening the library by DNA hybridization was provided by Clonetech (Palo Alto , California). From autoradiograms of duplicate filters, 11 plaques which hybridized to a corn CS cDNA probe were identified. After a second round of screening two of the original plaques, number 6-1 and number 10-1, showed positive hybridization. These plaques were tested with the probe a third time; and well isolated plaques were picked from each original. Following a fourth probing all the plaques hybridized, indicating that pure clones had been isolated.

DNA was prepared from these two phage clones, 6-1 and 10-1, using the protocol for plate lysate method [see Sambrook et al. (1989) *Molecular Cloning*: A Laboratory Manual, Cold Spring Harbor Laboratory Press]. Restriction endonuclease digests and agarose gel electrophoresis showed the two clones to be identical. The DNA fragments from the agarose gel were "Southern-blotted" [see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press] onto nylon filters and probed with labeled corn CS cDNA. A single 7.5 kb Sal I fragment and two Xba I fragments of 3.6 kb and 3.2 kb hybridized to the probe. The 3.2 kb Xba I fragment hybridized weakly to the probe whereas the 3.6 kb Xba I and the 7.5 kb Sal I fragments hybridized strongly.

The 3.6 kb Xba I fragment was cloned into the Xba I site of pGEM®-9Zf(-) that had been treated with calf intestinal alkaline phosphatase. Two subclones from each Xba I fragment with the fragments in both orientations with respect to pGEM®-9Zf(-) DNA were obtained following transformation of *E. coli*. The two 3.6 kb Xba I subclones were designated FS1179 and FS1180.

Restriction enzyme analysis of the subclones indicated that the 3.6 kb Xba I fragment in FS1179 and FS1180 included the 5' region of the corn CS gene. DNA from FS1180 was

sent to LARK Sequencing Technologies Inc. (Houston, TX) for complete DNA sequencing analysis. The sequence of the entire 3639 bp Xba I fragment is shown in see SEQ ID NO:18.

Complete sequence analysis of the 3639 bp Xba I fragment revealed 806 bp of sequence upstream from the protein coding region and 2833 bp of DNA encoding two-thirds of the corn CS protein. The 2833 bp includes seven exons and seven introns with the 3' Xba I site located in the seventh intron. Table 1 describes the location and length of exons and introns in the sequence as well the number of amino acids encoded by the exons. The first exon includes the entire chloroplast targeting signal and 12 amino acids into the region that shows amino acid sequence alignment with the *E. coli* protein. The last codon in Exon 7 encodes amino acid 333 of corn CS as shown in SEQ ID NO:17.

TABLE 5

REGION	FROM (bp)	TO (bp)	Length (bp)	Number of Encoded Amino Acids
Promoter	1	806	806	na*
Exon1	807	1194	387	129
Intron1	1195	1301	106	na
Exon2	1302	1405	103	35
Intron2	1406	1489	83	na
Exon3	1490	1563	73	24
Intron3	1564	1646	82	na
Exon4	1647	1815	168	57
Intron4	1816	2507	691	na
Exon5	2508	2567	59	20
Intron5	2568	2660	92	na
Exon6	2661	2864	203	68
Intron6	2865	2947	82	na
Exon7	2948	3034	86	29
Intron7	3035	3639	>604	na

*na = not applicable

Comparison of the corn CS cDNA sequence to the genomic CS DNA sequence indicated that the cDNA of clone FS1088 did not contain the entire chloroplast targeting signal as anticipated. The cDNA was not truncated on the 5' end, but instead contained a 170 bp deletion in the chloroplast transit sequence.

The complete amino acid sequence of the corn CS protein derived from combining the amino terminal sequence deduced from the corn genomic DNA fragment of SEQ ID

NO:18 and the carboxy terminal sequence from the corn cDNA fragment of SEQ ID NO:16 is shown in SEQ ID NO:19.

EXAMPLE 4

Modification of the Corn CS Gene and High level expression in *E. coli*

As indicated in Example 2, the open reading frame in plasmid FS1088 for the corn CS gene does not include an ATG initiator codon. Oligonucleotide adaptors OTG145 and OTG146 were designed to add an initiator codon in frame with the CS coding sequence.

OTG145	5'-AATTCATGAG TGCA-3'	SEQ ID NO:20
OTG146	5'-AATTTGCACT CATG-3'	SEQ ID NO:21

When annealed the oligonucleotides possess EcoR I sticky ends. Upon insertion into FS1088 in the desired orientation, an EcoR I site is present at the 5' end of the adaptor, the ATG initiator codon is within a BspH I restriction endonuclease site, and the EcoR I site at the 3' end of the adaptor is destroyed. The oligonucleotides were ligated into EcoR I digested FS1088, and insertion of the correct sequence in the desired orientation was verified by DNA sequencing.

To achieve high level expression of the corn CS gene in *E. coli* the bacterial expression vector pBT430 (see Example 2) was used. The corn CS gene was cut out of the modified FS1088 plasmid described above as an 1482 bp BspH I fragment and inserted into the expression vector pBT430 digested with Nco I. Clones with the CS gene in the proper orientation were identified by restriction enzyme mapping.

For high level expression each of the plasmids was transformed into *E. coli* strain BL21(DE3) or BL21(DE3)lysS [Studier et al., *J. Mol. Biol.*, (1986), *189*, 113-130]. Cultures were grown in LB medium containing ampicillin (100 mg/L) at 37°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) was added to a final concentration of 0.4 mM and incubation was continued overnight. The cells were collected by centrifugation and resuspended in 1/20th the original culture volume in 50 mM NaCl; 50 mM Tris-Cl, pH 7.5; 1 mM EDTA, and frozen at -20°C. Frozen aliquots of 1 mL were thawed at 37°C and sonicated, in an ice-water bath, to lyse the cells. The lysate was centrifuged at 4°C for 5 min at 12,000 rpm. The supernatant was removed and the pellet was resuspended in 1 mL of the above buffer.

The supernatant and pellet fractions of uninduced and IPTG-induced cultures were analyzed by SDS polyacrylamide gel electrophoresis. The best of the conditions tested was the induced culture of the BL21(DE3)lysS host. The major protein visible by Coomassie blue staining in the pellet fraction of this induced culture had a molecular weight of about 54 kd, the expected size for corn CS.

EXAMPLE 5

Isolation of the *E. coli* lysC Gene and mutations in lysC resulting in lysine-insensitive AKIII

The *E. coli* lysC gene has been cloned, restriction endonuclease mapped and sequenced previously [Cassan et al., *J. Biol. Chem.*, (1986), *261*, 1052-1057]. For the present invention the lysC gene was obtained on a bacteriophage lambda clone from an ordered library of 3400 overlapping segments of cloned *E. coli* DNA constructed by Kohara, Akiyama and Isono [Kohara et al., *Cell*, (1987), *50*, 595-508]. This library provides a physical map of the whole *E. coli* chromosome and ties the physical map to the genetic map. From the knowledge of the map position of lysC at 90 min. on the *E. coli* genetic map [Theze et al., *J. Bacteriol.*, (1974), *117*, 133-143], the restriction endonuclease map of the cloned gene [Cassan et al., *J. Biol. Chem.*, (1986), *261*, 1052-1057], and the restriction endonuclease map of the cloned DNA fragments in the *E. coli* library [Kohara et al., *Cell*, (1987), *50*, 595-508], it was possible to choose lambda phages 4E5 and 7A4 [Kohara et al., *Cell*, (1987), *50*, 595-508] as likely candidates for carrying the lysC gene. The phages were grown in liquid culture from single plaques as described [see Current Protocols in Molecular Biology (1987) Ausubel et al. eds. John Wiley & Sons New York] using LE392 as host [see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press]. Phage DNA was prepared by phenol extraction as described [see Current Protocols in Molecular Biology (1987) Ausubel et al. eds. John Wiley & Sons New York].

From the sequence of the gene several restriction endonuclease fragments diagnostic for the lysC gene were predicted, including an 1860 bp EcoR I-Nhe I fragment, a 2140 bp EcoR I-Xmn I fragment and a 1600 bp EcoR I-BamH I fragment. Each of these fragments was detected in both of the phage DNAs confirming that these carried the lysC gene. The EcoR I-Nhe I fragment was isolated and subcloned in plasmid pBR322 digested with the same enzymes, yielding an ampicillin-resistant, tetracycline-sensitive *E. coli* transformant. The plasmid was designated pBT436.

To establish that the cloned lysC gene was functional, pBT436 was transformed into *E. coli* strain Gif106M1 (*E. coli* Genetic Stock Center strain CGSC-5074) which has mutations in each of the three *E. coli* AK genes [Theze et al., *J. Bacteriol.*, (1974), *117*, 133-143]. This strain lacks all AK activity and therefore requires diaminopimelate (a precursor to lysine which is also essential for cell wall biosynthesis), threonine and methionine. In the transformed strain all these nutritional requirements were relieved demonstrating that the cloned lysC gene encoded functional AKIII.

Addition of lysine (or diaminopimelate which is readily converted to lysine in vivo) at a concentration of approximately 0.2 mM to the growth medium inhibits the growth of Gif106M1 transformed with pBT436. M9 media [see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press] supplemented with

the arginine and isoleucine, required for Gif106M1 growth, and ampicillin, to maintain selection for the pBT436 plasmid, was used. This inhibition is reversed by addition of threonine plus methionine to the growth media. These results indicated that AKIII could be inhibited by exogenously added lysine leading to starvation for the other amino acids derived from aspartate. This property of pBT436-transformed Gif106M1 was used to select for mutations in lysC that encoded lysine-insensitive AKIII.

Single colonies of Gif106M1 transformed with pBT436 were picked and resuspended in 200 μL of a mixture of 100 μL 1% lysine plus 100 μL of M9 media. The entire cell suspension containing 10^7-10^8 cells was spread on a petri dish containing M9 media supplemented with the arginine, isoleucine, and ampicillin. Sixteen petri dishes were thus prepared. From 1 to 20 colonies appeared on 11 of the 16 petri dishes. One or two (if available) colonies were picked and retested for lysine resistance and from this nine lysine-resistant clones were obtained. Plasmid DNA was prepared from eight of these and re-transformed into Gif106M1 to determine whether the lysine resistance determinant was plasmid-borne. Six of the eight plasmid DNAs yielded lysine-resistant colonies. Three of these six carried lysC genes encoding AKIII that was uninhibited by 15 mM lysine, whereas wild type AKIII is 50% inhibited by 0.3-0.4 mM lysine and >90% inhibited by 1 mM lysine.

To determine the molecular basis for lysine-resistance the sequences of the wild type lysC gene and three mutant genes were determined. The sequence of the wild type lysC gene cloned in pBT436 (SEQ ID NO:22) differed from the published lysC sequence in the coding region at 5 positions. Four of these nucleotide differences were at the third position in a codon and would not result in a change in the amino acid sequence of the AKIII protein. One of the differences would result in a cysteine to glycine substitution at amino acid 58 of AKIII. These differences are probably due to the different strains from which the lysC genes were cloned.

The sequences of the three mutant lysC genes that encoded lysine-insensitive AK each differed from the wild type sequence by a single nucleotide, resulting in a single amino acid substitution in the protein. Mutant M2 had an A substituted for a G at nucleotide 954 of SEQ ID NO:22 resulting in an isoleucine for methionine substitution at amino acid 318 and mutants M3 and M4 had identical T for C substitutions at nucleotide 1055 of SEQ ID NO:22 resulting in an isoleucine for threonine substitution at amino acid 352. Thus, either of these single amino acid substitutions is sufficient to render the AKIII enzyme insensitive to lysine inhibition.

An Nco I (CCATGG) site was inserted at the translation initiation codon of the lysC gene using the following oligonucleotides:

5'-GATCCATGGC TGAAATTGTT GTCTCCAAAT TTGGCG-3' SEQ ID NO:24

5'-GTACCGCCAA ATTTGGAGAC AACAATTTCA GCCATG-3' SEQ ID NO:25

When annealed these oligonucleotides have BamH I and Asp 718 "sticky" ends. The plasmid pBT436 was digested with BamH I, which cuts upstream of the lysC coding sequence and Asp 718 which cuts 31 nucleotides downstream of the initiation codon. The annealed oligonucleotides were ligated to the plasmid vector and *E. coli* transformants were obtained. Plasmid DNA was prepared and screened for insertion of the oligonucleotides based on the presence of an Nco I site. A plasmid containing the site was sequenced to assure that the insertion was correct, and was designated pBT457. In addition to creating an Nco I site at the initiation codon of lysC, this oligonucleotide insertion changed the second codon from TCT, coding for serine, to GCT, coding for alanine. This amino acid substitution has no apparent effect on the AKIII enzyme activity.

The lysC gene was cut out of plasmid pBT457 as a 1560 bp Nco I-EcoR I fragment and inserted into the expression vector pBT430 digested with the same enzymes, yielding plasmid pBT461. For expression of the mutant lysC-M4 gene pBT461 was digested with Kpn I-EcoR I, which removes the wild type lysC gene from about 30 nucleotides downstream from the translation start codon, and inserting the analogous Kpn I-EcoR I fragments from the mutant genes yielding plasmid pBT492.

EXAMPLE 6

Molecular Cloning of Corn Genes Encoding Methionine-Rich Seed Storage Proteins

A high methionine 10 kD zein gene [Kirihara et al., *Mol. Gen. Genet.*, (1988), *211*, 477-484] was isolated from corn genomic DNA using PCR. Two oligonucleotides 30 bases long flanking this gene were synthesized using an Applied Biosystems DNA synthesizer. Oligomer SM56 (SEQ ID NO:26) codes for the positive strand spanning the first ten amino acids:

SM56 5'-ATGGCAGCCA AGATGCTTGC ATTGTTCGCT-3' SEQ ID NO:26

Oligomer CFC77 (SEQ ID NO:27) codes for the negative strand spanning the last ten amino acids:

CFC77 5'-GAATGCAGCA CCAACAAAGG GTTGCTGTAA-3' SEQ ID NO:27

These were employed to generate by polymerase chain reaction (PCR) the 10 kD coding region using maize genomic DNA from strain B85 as the template. PCR was performed using a Perkin-Elmer Cetus kit according to the instructions of the vendor on a thermocycler manufactured by the same company. The reaction product, when run on a 1% agarose gel and stained with ethidium bromide, showed a strong DNA band of the size

expected for the 10 kD zein gene, 450 bp, with a faint band at about 650 bp. The 450 bp band was electro-eluted onto DEAE cellulose membrane (Schleicher & Schuell) and subsequently eluted from the membrane at 65°C with 1 M NaCl, 0.1 mM EDTA, 20 mM Tris-Cl, pH 8.0. The DNA was ethanol precipitated and rinsed with 70% ethanol and dried. The dried pellet was resuspended in 10 µL water and an aliquot (usually 1 µL) was used for another set of PCR reactions, to generate by asymmetric priming single-stranded linear DNAs. For this, the primers SM56 and CFC77 were present in a 1:20 molar ratio and 20:1 molar ratio. The products, both positive and negative strands of the 10 kD zein gene, were phenol extracted, ethanol precipitated, and passed through NACS (Bethesda Research Laboratories) columns to remove the excess oligomers. The eluates were ethanol precipitated twice, rinsed with 70% ethanol, and dried. DNA sequencing was done using the appropriate complementary primers and a sequenase kit from United States Biochemicals Company according to the vendors instructions. The sequence deviated from the published coding sequence (Kirihara et al., *Gene*, (1988), *71*, 359-370) in one base pair at nucleotide position 1504 of the published sequence. An A was changed to a G which resulted in the change of amino acid 123 (with the initiator methionine as amino acid 1) from Gln to Arg. It is not known if the detected mutation was generated during the PCR reaction or if this is another allele of the maize 10 kD zein gene. A radioactive probe was made by nick-translation of the PCR-generated 10 kD zein gene using ^{32}P-dCTP and a nick-translation kit purchased from Bethesda Research Laboratories.

A genomic library of corn in bacteriophage lambda was purchased from Clontech (Palo Alto, CA). Data sheets from the supplier indicated that the corn DNA was from seven-day-old seedlings grown in the dark. The vector was l-EMBL-3 carrying BamHI fragments 15 kb in average size. A titer of 1 to 9 x 10^9 plaque forming units (pfu)/mL was indicated by the supplier. Upon its arrival the library was titered and contained 2.5 x 10^9 pfu/mL.

The protocol for screening the library by DNA hybridization was provided by the vendor. About 30,000 pfu were plated per 150-mm plate on a total of 15 Luria Broth (LB) agar plates giving 450,000 plaques. Plating was done using *E. coli* LE392 grown in LB + 0.2% maltose as the host and LB-7.2% agarose as the plating medium. The plaques were absorbed onto nitrocellulose filters (Millipore HATF, 0.45 mM pore size), denatured in 0.5M NaOH, neutralized in 1.5 M NaCl, 0.5 M Tris-Cl pH 7.5, and rinsed in 3XSSC [Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press]. The filters were blotted on Whatman 3MM paper and heated in a vacuum oven at 80°C for two hours to allow firm anchorage of phage DNA in the membranes.

The ^{32}P-labelled 10 kD zein was used as a hybridization probe to screen the library. The fifteen 150-mm nitrocellulose filters carrying the l phage plaques were screened using radioactive 10 kD gene probe. After four hours prehybridizing at 60°C in 50XSSPE, 5X Denhardt's, [see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold

Spring Harbor Laboratory Press] 0.1% SDS, 100 µg/mL calf thymus DNA, the filters were transferred to fresh hybridization mix containing the denatured radiolabeled 10 kD zein gene (cpm/mL) and stored overnight at 60°C. They were rinsed the following day under stringent conditions: one hour at room temp in 2XSSC - 0.05% SDS and one hour at 68°C in 1XSSC - 0.1% SDS. Blotting on 3MM Whatman paper followed, then air drying and autoradiography at -70°C with Kodak XAR-5 films with DuPont Cronex® Lightning Plus intensifying screens. From these autoradiograms, 20 hybridizing plaques were identified. These plaques were picked from the original petri plate and plated out at a dilution to yield about 100 plaques per 80-mm plate. These plaques were absorbed to nitrocellulose filters and re-probed using the same procedure. After autoradiography only one of the original plaques, number 10, showed two hybridizing plaques. These plaques were tested with the probe a third time; all the progeny plaques hybridized, indicating that pure clones had been isolated.

DNA was prepared from these two phage clones, 10-1, 10-2, using the protocol for DNA isolation from small-scale liquid l-phage lysates (Ansul et al. (1987) Current Protocols in Molecular Biology, pp. 1.12.2, 1.13.5-6). Restriction endonuclease digests and agarose gel electrophoresis showed the two clones to be identical. The DNA fragments from the agarose gel were "Southern-blotted" [see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press] onto nitrocellulose membrane filters and probed with radioactively-labeled 10 kD zein DNA generated by nick translation. A single 7.5 kb BamH I fragment and a single 1.4 kb Xba I fragment hybridized to the probe.

The 7.5 kb BamH I fragment was isolated from a BamH I digest of the l DNA run on an 0.5% low melting point (LMP) agarose gel. The 7.5 kb band was excised, melted, and diluted into 0.5 M NaCl and loaded onto a NACS column, which was then washed with 0.5 M NaCl, 10 mM Tris-Cl, pH 7.2, 1 mM EDTA and the fragment eluted with 2 M NaCl, 10 mM Tris-Cl, pH 7.2, 1 mM EDTA. This fragment was ligated to the phagemid pTZ18R (Pharmacia) which had been cleaved with BamH I and treated with calf intestinal alkaline phosphatase [see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press] to prevent ligation of the phagemid to itself. Subclones with these fragments in both orientations with respect to the pTZ18R DNA were obtained following transformation of *E. coli*.

An Xba I digest of the cloned l phage DNA was run on an 0.8% agarose gel and a 1.4 kb fragment was isolated using DEAE cellulose membrane (same procedure as for the PCR-generated 10 kD zein DNA fragment described above). This fragment was ligated to pTZ18R cut with Xba I in the same way as described above. Subclones with these fragments in both orientations with respect to the pTZ18R DNA, designated pX8 and pX10, were obtained following transformation of *E. coli*. Single-stranded DNAs were made from the subclones using the protocol provided by Pharmacia. The entire 1.4 kb Xba I fragments

were sequenced. An additional 700 bases adjacent to the Xba I fragment was sequenced from the BamH I fragment in clone pB3 (fragment pB3 is in the same orientation as pX8) giving a total of 2123 bases of sequence (SEQ ID NO:28).

Encoded on this fragment is another methionine-rich zein, which is related to the 10 kD zein and has been designated High Sulfur Zein (HSZ) [see World Patent Publication No. WO 92/14822]. From the deduced amino acid sequence of the protein, its molecular weight is approximately 21 kD and it is about 38% methionine by weight.

EXAMPLE 7

Modification of the HSZ Gene by Site-Directed Mutagenesis

Three Nco I sites were present in the 1.4 kD Xba I fragment carrying the HSZ gene, all in the HSZ coding region. It was desirable to maintain only one of these sites (nucleotides 751-756 in SEQ ID NO:28) that included the translation start codon. Therefore, the Nco I sites at positions 870-875 and 1333-1338 were eliminated by oligonucleotide-directed site-specific mutagenesis [see Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press]. The oligonucleotides synthesized for the mutagenesis were:

CFC99	5'-ATGAACCCTT GGATGCA-3'	SEQ ID NO:30
CFC98	5'-CCCACAGCAA TGGCGAT-3'	SEQ ID NO:31

Mutagenesis was carried out using a kit purchased from Bio-Rad (Richmond, CA), following the protocol provided by the vendor.

The process changed the A to T at 872 and the C to A at 1334. These were both at the third position of their respective codons and resulted in no change in the amino acid sequence encoded by the gene, with C C A to C C T, still coding for Pro and G C C to G C A, still coding for Ala. The plasmid clone containing the modified HSZ gene with a single Nco I site at the ATG start codon was designated pX8m. Because the native HSZ gene has a unique Xba I site at the stop codon of the gene (1384-1389, SEQ ID NO:28), a complete digest of the DNA with Nco I and Xba I yields a 637 bp fragment containing the entire coding sequence of the precursor HSZ polypeptide (SEQ ID NO:32).

It was desirable to create a form of the HSZ gene with alternative unique restriction endonuclease sites just past the end of the coding region. To do this oligonucleotides CFC104 (SEQ ID NO:34) and CFC105 (SEQ ID NO:35):

CFC104	5'-CTAGCCCGGGTAC -3'	(SEQ ID NO:34)
CFC105	3'- GGGCCCATGGATC-5'	(SEQ ID NO:35)

were annealed and ligated into the Xba I site, introducing two new restriction sites, Sma I and Kpn I, and destroying the Xba I site. The now unique Xba I site from nucleotide 1-6 in SEQ ID NO:28 and the Ssp I site from nucleotide 1823-1828 in SEQ ID NO:28 were used to obtain a fragment that included the HSZ coding region plus its 5' and 3' regulatory regions. This fragment was cloned into the commercially-available vector pTZ19R (Pharmacia) digested with Xba I and Sma I, yielding plasmid pCC10.

It was desirable to create an altered form of the HSZ gene with a unique restriction endonuclease site at the start of the mature protein, i.e., with the amino terminal signal sequence removed. To accomplish this a DNA fragment was generated using PCR. Template DNA for the PCR reaction was plasmid pX8m. Oligonucleotide primers for the reaction were:

CFC106: 5'-CCACTTCATGACCCATATCCCAGGGCACTT-3' SEQ ID NO:36

CFC88: 5'-TTCTATCTAGAATGCAGCACCAACAAAGGG-3' SEQ ID NO:37

The CFC106 (SEQ ID NO:36) oligonucleotide provided the PCR-generated fragment with a BspH I site (underlined), which when digested with BspH I results in a cohesive-end identical to that generated by an Nco I digest. This site was located at the junction of the signal sequence and the mature HSZ coding sequence. The CFC88 (SEQ ID NO:37) oligonucleotide provided the PCR-generated fragment with an Xba I site (underlined) at the translation terminus of the HSZ gene. The BspH I-Xba I fragment (SEQ ID NO:38) obtained by digestion of the PCR-generated fragment, encodes the mature form of HSZ with the addition of a methionine residue at the amino terminus of the protein to permit initiation of translation.

EXAMPLE 8

Construction of Chimeric Genes for Expression of CS, AKIII-M4, and HSZ Proteins in the Seeds of Monocot Plants

The following chimeric genes were made for transformation into monocot plants:

globulin 1 promoter/mcts/lysC-M4/globulin 1 3' region
globulin 1 promoter/corn CS coding region/globulin 1 3' region
glutelin 2 promoter/mcts/lysC-M4/NOS 3' region
glutelin 2 promoter/corn CS coding region/10 kD 3' region
10 kD promoter/HSZ coding region/10 kD 3' region
glutelin 2 promoter/HSZ coding region/10 kD 3' region

A gene expression cassette employing the 10 kD zein regulatory sequences includes about 925 nucleotides upstream (5') from the translation initiation codon and about 945

nucleotides downstream (3') from the translation stop codon. The entire cassette is flanked by an EcoR I site at the 5' end and BamH I, Sal I and Hind III sites at the 3' end. The DNA sequence of these regulatory regions have been described in the literature [Kirihara et al,. *Gene*, (1988), *71*, 359-370] and DNA fragments carrying these regulatory sequences were obtained from corn genomic DNA via PCR. Between the 5' and 3' regions is a unique Nco I site, which includes the ATG translation initiation codon. The oligonucleotides CFC104 (SEQ ID NO:34) and CFC105 (SEQ ID NO:35) (see Example 7) were inserted at the Xba I site near the 10 kD zein translation stop codon, thus adding a unique Sma I site. An Nco I-Sma I fragment containing the HSZ coding region was isolated from plasmid pCC10 (see Example 7) and inserted into Nco I-Sma I digested 10 kD zein expression cassette creating the chimeric gene: 10 kD promoter/HSZ coding region/10 kD 3' region.

The glutelin 2 promoter was cloned from corn genomic DNA using PCR with primers based on the published sequence [Reina et al., *Nucleic Acids Res.*, (1990), *18*, 6426-6426]. The promoter fragment includes 1020 nucleotides upstream from the ATG translation start codon. An Nco I site was introduced via PCR at the ATG start site to allow for direct translational fusions. A BamH I site was introduced on the 5' end of the promoter. The 1.02 kb BamH I to Nco I promoter fragment was linked to an Nco I to Hind III fragment carrying the HSZ coding region/10 kD 3' region described above yielding the chimeric gene: glutelin 2 promoter/HSZ coding region/10 kD 3' region in a plasmid designated pML103.

The globulin 1 promoter and 3' sequences were isolated from a Clontech corn genomic DNA library using oligonucleotide probes based on the published sequence of the globulin 1 gene [Kriz et al., *Plant Physiol.*, (1989), *91*, 636]. The cloned segment includes the promoter fragment extending 1078 nucleotides upstream from the ATG translation start codon, the entire globulin coding sequence including introns and the 3' sequence extending 803 bases from the translational stop. To allow replacement of the globulin 1 coding sequence with other coding sequences an Nco I site was introduced at the ATG start codon, and Kpn I and Xba I sites were introduced following the translational stop codon via PCR to create vector pCC50. There is a second Nco I site within the globulin 1 promoter fragment. The globulin 1 gene cassette is flanked by Hind III sites.

Plant amino acid biosynthetic enzymes are known to be localized in the chloroplasts and therefore are synthesized with a chloroplast targeting signal. Bacterial proteins such as AKIII have no such signal. A chloroplast transit sequence (cts) was therefore fused to the lysC-M4 coding sequence in the chimeric genes described below. For corn the cts used was based on the the cts of the small subunit of ribulose 1,5-bisphosphate carboxylase from corn [Lebrun et al., *Nucleic Acids Res.*, (1987), *15*, 4360] and is designated mcts. The oligonucleotides SEQ ID NOS:40-45 were synthesized and used to attach the mcts to lysC-M4.

Oligonucleotides SEQ ID NO:40 and SEQ ID NO:41, which encode the carboxy terminal part of the corn chloroplast targeting signal, were annealed, resulting in Xba I and Nco I compatible ends, purified via polyacrylamide gel electrophoresis, and inserted into Xba I plus Nco I digested pBT492 (see Example 5). The insertion of the correct sequence was verified by DNA sequencing yielding pBT556. Oligonucleotides SEQ ID NO:42 and SEQ ID NO:43, which encode the middle part of the chloroplast targeting signal, were annealed, resulting in Bgl II and Xba I compatible ends, purified via polyacrylamide gel electrophoresis, and inserted into Bgl II and Xba I digested pBT556. The insertion of the correct sequence was verified by DNA sequencing yielding pBT557. Oligonucleotides SEQ ID NO:44 and SEQ ID NO:45, which encode the amino terminal part of the chloroplast targeting signal, were annealed, resulting in Nco I and Afl II compatible ends, purified via polyacrylamide gel electrophoresis, and inserted into Nco I and Afl II digested pBT557. The insertion of the correct sequence was verified by DNA sequencing yielding pBT558. Thus the mcts was fused to the lysC-M4 gene.

To construct the chimeric gene: globulin 1 promoter/mcts/lysC-M4/globulin 1 3' region an Nco I to Hpa I fragment containing the mcts/lysC-M4 coding sequence was isolated from plasmid pBT558 and inserted into Nco I plus Sma I digested pCC50 creating plasmid pBT663.

To construct the chimeric gene: glutelin 2 promoter/mcts/lysC-M4/NOS 3' region the 1.02 kb BamH I to Nco I glutelin 2 promoter fragment described above was linked to the Nco I to Hpa I fragment containing the mcts/lysC-M4 coding sequence described above and to a Sma I to Hind III fragment carrying the NOS 3' region creating.

To construct the chimeric gene: globulin 1 promoter/corn CS coding region/globulin 1 3' region a 1482 base pair BspH I fragment containing the corn CS coding region (see Example 4) was isolated and inserted into an Nco I partial digest of pCC50. A plasmid designated pML157 carried the CS coding region in the proper orientation to create the indicated chimeric gene, as determined via restriction endonuclease digests.

To construct the chimeric gene: glutelin 2 promoter/corn CS coding region/10 kD 3' region the HSZ coding region was removed from pML103 (above) by digestion with Nco I and Xma I and insertion of an oligonucleotide adaptor containing an EcoR I site and Nco I and Xma I sticky ends. The resulting plasmid was digested with Nco I and the 1482 base pair BspH I fragment containing the corn CS coding region (see above and Example 4) was inserted. A plasmid with the CS coding region in the proper orientation, as determined via restriction endonuclease digests, was obtained, creating the indicated chimeric gene.

A corn CS gene that contained the entire chloroplast targeting signal was constructed by fusing the 5' end of the genomic CS gene to the 3' end of the cDNA. A 697 bp Nco I to Sph I genomic DNA fragment replaced the analogous Nco I to Sph I fragment in the cDNA. Thus, the first 168 amino acids are encoded by the genomic CS sequence and the coding

sequence is interrupted by two introns. The remaining 341 amino acids are encoded by cDNA CS sequence with no further introns, resulting in a protein of 509 amino acids in length (SEQ ID NO:19). A 1750 bp Nco I to BspH I DNA fragment that includes the entire CS coding region was inserted into the corn embryo and endosperm expression cassettes resulting in the chimeric genes globulin 1 promoter/corn CS coding region/globulin 1 3' region in plasmid pFS1198 and glutelin 2 promoter/corn CS coding region/10 kD zein 3' region in plasmid pFS1196, respectively.

EXAMPLE 9

Construction of Chimeric Genes for Expression of CS, AKIII-M4, and HSZ Proteins in the Seeds of Dicot Plants

The following chimeric genes were made for transformation into dicot plants:

phaseolin promoter/scts/lysC-M4/phaseolin 3' region
KTI3 promoter/scts/corn CS coding region/KTI3 3' region
phaeolin promoter/HSZ coding region/phaseolin 3' region
β–conglycinin-conglycinin promoter/HSZ coding region/phaseolin 3' region

A first seed-specific expression cassette used for expression in dicotyledonous plants is composed of the promoter and transcription terminator from the gene encoding the b subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* [Doyle et al. (1986) J. Biol. Chem. 261:9228-9238]. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

A second seed-specific expression cassette used for expression in dicotyledonous plants is composed of the promoter from the α´ subunit of soybean β-conglycinin (*Glycine max*) and the transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (above). The conglycinin cassette includes 607 nucleotides upstream (5') from the translation initiation codon of soybean β-conglycinin and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

A third seed-specific expression cassette used for expression in dicotyledonous plants is composed of the promoter and transcription terminator from the soybean Kunitz trysin inhibitor 3 (KTI3) gene [Jofuku et al., *Plant Cell*, (1989), *1*, 427-435]. The KTI3 cassette includes about 2000 nucleotides upstream (5') from the translation initiation codon and about

240 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco 1 (which includes the ATG translation initiation codon), Xba I, Kpn 1 and Sma 1. The entire cassette is flanked by BamH I sites.

Plant amino acid biosynthetic enzymes are known to be localized in the chloroplasts and therefore are synthesized with a chloroplast targeting signal. Bacterial proteins such as AKIII have no such signal. A chloroplast transit sequence (cts) was therefore fused to the lysC-M4 coding sequence in some chimeric genes. The cts used was based on the the cts of the small subunit of ribulose 1,5-bisphosphate carboxylase from soybean [Berry-Lowe et al., *J. Mol. Appl. Genet.*, (1982), *1*, 483-498]. The oligonucleotides SEQ ID NOS:46-51 were synthesized and used as described below. The soybean cts (scts) was also used to replace the native corn cts in the corn CS gene.

Oligonucleotides SEQ ID NO:46 and SEQ ID NO:47, which encode the carboxy terminal part of the chloroplast targeting signal, were annealed, resulting in Nco I compatible ends, purified via polyacrylamide gel electrophoresis, and inserted into Nco I digested pBT461. The insertion of the correct sequence in the correct orientation was verified by DNA sequencing yielding pBT496. Oligonucleotides SEQ ID NO:48 and SEQ ID NO:49, which encode the amino terminal part of the chloroplast targeting signal, were annealed, resulting in Nco I compatible ends, purified via polyacrylamide gel electrophoresis, and inserted into Nco I digested pBT496. The insertion of the correct sequence in the correct orientation was verified by DNA sequencing yielding pBT521. Thus the scts was fused to the lysC gene.

To fuse the scts to the lysC-M4 gene, pBT521 was digested with Sal I, and an approximately 900 bp DNA fragment that included the scts and the amino terminal coding region of lysC was isolated. This fragment was inserted into Sal I digested pBT492, effectively replacing the amino terminal coding region of lysC-M4 with the fused scts and the amino terminal coding region of lysC. Since the mutation that resulted in lysine-insensitivity was not in the replaced fragment, the new plasmid, pBT523, carried the scts fused to lysC-M4.

A 1600 bp Nco I-Hpa I fragment containing the cts fused to lysC-M4 plus about 90 bp of 3' non-coding sequence was isolated from pBT523 and inserted into the phaseolin seed-specific expression cassette digested with Nco I and Sma I, yielding plasmid pBT544 carrying the chimeric gene:

phaseolin promoter/scts/lysC-M4/phaseolin 3' region.

An scts DNA fragment that can be readily inserted into dicot gene expression cassettes was created. Employing PCR with primers CF32 (SEQ ID NO:50) and CF33 (SEQ ID NO:51) and any template DNA carrying the soybean cts, e.g. pBT523 above, results in a

DNA fragment carrying the entire scts. .This fragment is then cut with Nco I and ligated to any gene that carries an Nco I site in-frame with the translation initiation codon.

The corn CS gene of plasmid pFS1088 (Example 3) was cut with restriction enzyme Sst II and the oligonucleotide adaptor shown in SEQ ID NO:52 was self-annealed and inserted. This removes most of the corn chloroplast transit peptide coding region and adds an Nco I site in-frame with the CS coding sequence. A DNA fragment containing the thus modified corn CS gene was was obtained by digestion with Nco I and BspH I and ligated into the KTI3 expression cassette digested with Nco I. Insertion of the corn CS gene in the proper orientation was determined by restriction enzyme mapping. The scts was then added as an Nco I fragment as described above yielding the chimeric gene: KTI3 promoter/scts/corn CS coding region/KTI3 3' region.

The Nco I-Xba I fragment containing the entire HSZ coding region (see Example 7) was isolated from an agarose gel following electrophoresis and inserted into the phaseolin and β-conglycinin expression cassettes which had been digested with Nco I-Xba I. Thus the two chimeric genes:

1) phaseolin 5' region/HSZ/phaseolin 3' region
2) β-conglycinin 5' region/HSZ/phaseolin 3' region were created.

EXAMPLE 10

Isolation of the *E. coli* metL Gene and Construction of Chimeric Genes for Expression in the Seeds of Plants

The metL gene of *E. coli* encodes a bifunctional protein, AKII-HDHII; the AK and HDH activities of this enzyme are insensitive to all pathway end-products. The metL gene of *E. coli* has been isolated and sequenced previously [Zakin et al., *J. Biol. Chem.*, (1983), *258*, 3028-3031]. For the present invention a DNA fragment containing the metL gene was isolated and modified from *E. coli* genomic DNA obtained from strain LE392 using PCR. The following PCR primers were designed and synthesized:

CF23 5'-GAAACCATGG CCAGTGTGAT TGCGCAGGCA-3' SEQ ID NO:53

CF24 5'-GAAAGGTACC TTACAACAAC TGTGCCAGC-3' SEQ ID NO:54

These primers add an Nco I site which includes a translation initiation codon at the amino terminus of the AKII-HDHII protein. In order to add the restriction site and additional codon, GCC coding for alanine, was also added to the amino terminus of the protein. The primers also add a Kpn I site immediately following the translation stop codon.

PCR was performed using a Perkin-Elmer Cetus kit according to the instructions of the vendor on a thermocycler manufactured by the same company. The primers were at a concentration of 10 mM and the thermocycling conditions were:

94° 1 min, 50° 2 min, 72° 8 min for 10 cycles followed by
94° 1 min, 72° 8 min for 30 cycles.

Reactions with four different concentrations of template DNA all yielded the expected 2.4 kb DNA fragment, along with several other smaller fragments. The four PCR reaction mixes were pooled, digested with Nco I and Kpn I and the 2.4 kb fragments were purified and isolated from an agarose gel. The fragment was inserted into a modified pBT430 expression vector (see Example 2) containing a Kpn I site downstream of the Nco I site at the translation initiation codon. DNA was isolated from 8 clones carrying the 2.4 kb fragment in the pBT430 expression vector and transformed into the expression host strain BL21(DE3). Cultures were grown in TB medium containing ampicillin (100 mg/L) at 37°C overnight. The cells were collected by centrifugation and resuspended in 1/25th the original culture volume in 50 mM NaCl; 50 mM Tris-Cl, pH 7.5; 1 mM EDTA, and frozen at -20°C, thawed at 37°C and sonicated, in an ice-water bath, to lyse the cells. The lysate was centrifuged at 4°C for 5 min at 12,000 rpm. The supernatant was removed and the pellet was resuspended in the above buffer. The supernatant fractions were assayed for HDH enzyme activities to identify clones expressing functional proteins. HDH activity was assayed as shown below:

HDH ASSAY

Stock solutions	1.0 µl	0.20 µl	Final conc
0.2 M KPO_4, pH 7.0	500 µl	100 µl	100 mM
3.7 M KCl	270 µl	54 µl	1.0 M
0.5 M EDTA	20 µl	4 µl	10 mM
1.0 M $MgCl_2$	10 µl	2 µl	10 mM
2 mM NADPH	100 µl	20 µl	0.20 mM

Make Mixture of above reagents with amounts multiplied by number of assays.
Use 0.9 mL of mix for 1mL assay; 180 µL of mix for 0.2 mL assay in microtiter dish

Add			
1.0M ASA in 1.0N HCl	1µl	0.2µl	1.0mM
to 1/2 the assay mix; remaining 1/2 lacks ASA to serve as blank			
enzyme extract	10-100 µl	2-20 µl	
H_20	to 1.0 mL	to 0.20 mL	

Add enzyme extract last to start reaction. Incubate at ~30°C; monitor NADPH oxidation at 340 nM. 1 unit oxidizes 1 µmol NADPH/min at 30°C in the 1 mL reaction.

Four of eight extracts showed HDH activity well above the control. These four were then assayed for AK activity. AK activity was assayed as shown below:

AK ASSAY

Assay mix (for 12 X 1.0mL or 48 X 0.25mL assays):

2.5 mL H_20
2.0 mL 4M KOH
2.0 mL 4M NH_2OH-HCl
1.0 mL 1M Tris-HCl pH 8.0
0.5 mL 0.2M ATP (121 mg/ml in 0.2M NaOH)
50 µL 1M $MgSO_4$

pH of assay mix should be 7-8

Each 1.5 mL eppendorf assay tube contains:

	MACRO assay	micro assay
assay mix	0.64 mL	0.16 mL
0.2M L-Aspartate	0.04 mL	0.01 mL
extract	5-120 µL	1-30 µl
H_20 to total vol.	0.8 mL	0.2 mL
Assay tubes are incubated at 30°C for 30-60 min		
Add to develop color;		
$FeCl_3$ reagent	0.4 mL	0.1 mL
$FeCl_3$ reagent is:	10% w/v $FeCl_3$	50 g
	3.3% TCA	15.5 g
	0.7% HCl	35 mL HCl
		H_2O to 500 mL

Spin for 2 min in eppendorf centrifuge tube.
Read OD at 540 nm.

Two extracts also had high levels of AK enzyme activity. These two extracts were then tested for inhibition of AK or HDH activity by the pathway end-products, lys, thr and met. Neither the AK nor the HDH activity of the extract from clone 5 was inhibited by 30 mM concentrations of any of the end-products.

The supernatant and pellet fractions of several of the extracts were also analyzed by SDS polyacrylamide gel electrophoresis. In the extract from clone 5, the major protein visible by Coomassie blue staining in both the pellet and supernatant fractions had a molecular weight of about 85 kd, the expected size for AKII-HDHII. The metL gene in plasmid pBT718 from clone 5 was used for all subsequent work.

Plant amino acid biosynthetic enzymes are known to be localized in the chloroplasts and therefore are synthesized with a chloroplast targeting signal. Bacterial proteins have no such signal. A chloroplast transit sequence (cts) was therefore fused to the metL coding

sequence in the chimeric genes described below. For corn the cts used was based on the the cts of the small subunit of ribulose 1,5-bisphosphate carboxylase from corn [Lebrun et al. (1987) *Nucleic Acids Res. 15*:4360] and is designated mcts.

Oligonucleotides SEQ ID NO:40 and SEQ ID NO:41, which encode the carboxy terminal part of the corn chloroplast targeting signal, were annealed, resulting in Xba I and Nco I compatible ends, purified via polyacrylamide gel electrophoresis, and inserted into Xba I plus Nco I digested pBT718. The insertion of the correct sequence was verified by DNA sequencing yielding pBT725. To complete the corn chloroplast targeting signal, pBT725 was digested with Bgl II and Xba I, and a 1.14 kb BamH I to Xba I fragment from pBT580 containing the glutelin 2 promoter plus the amino terminal part of the corn chloroplast targeting signal was inserted creating pBT726.

To construct the chimeric gene:

globulin 1 promoter/mcts/metL/globulin 1 3' region

the 2.6 kb Nco I to Kpn I fragment containing the mcts/metL coding sequence was isolated from plasmid pBT726 and inserted into Nco I plus Kpn I digested pCC50 creating plasmid pBT727.

To construct the chimeric gene:

glutelin 2 promoter/mcts/metL/NOS 3' region

the 2.6 kb Nco I to Kpn I fragment containing the mcts/metL coding sequence was isolated from plasmid pBT726 and linked to the 1.02 kb BamH I to Nco I glutelin 2 promoter fragment described in Example 8 and to a Kpn I to Hind III fragment carrying the NOS 3' region creating plasmid pBT728.

To construct the chimeric gene:

phaseolin promoter/scts/metL/phaseolin 3' region

the 2.4 kb Nco I to Kpn I fragment containing the metL coding sequence was isolated from plasmid pBT718 and inserted into Nco I plus Kpn I digested phaseolin expression cassette. The scts was then added as an Nco I fragment as described in Example 8.

EXAMPLE 11

Construction of Chimeric MS Genes for Expression in the Seeds of Plants

The following chimeric genes were made for transformation into monocot plants:

globulin 1 promoter/tobacco MS coding region/globulin 1 3' region;

glutelin 2 promoter/tobacco MS coding region/NOS 3' region.

To construct the chimeric gene:

globulin 1 promoter/tobacco MS coding region/globulin 1 3' region, the 2300 bp BspH I-Kpn I fragment containing the tobacco MS coding sequence was isolated from plasmid pBT773 and inserted into Nco I plus Kpn I digested pCC50 (Example 8).

To construct the chimeric gene:

glutelin 2 promoter/tobacco MS coding region/NOS 3' region, the 2300 bp BspH I-Kpn I fragment containing the tobacco MS coding sequence was isolated from plasmid pBT773 and linked to the 1.02 kb BamH I to Nco I glutelin 2 promoter fragment described in Example 8 and to a Kpn I to Hind III fragment carrying the NOS 3' region.

The following chimeric gene was made for transformation into dicot plants: KTI3 promoter/tobacco MS coding region/KTI3 3' region.

To construct the chimeric gene:

KTI3 promoter/tobacco MS coding region/KTI3 3' region, the 2300 bp BspH I-Kpn I fragment containing the tobacco MS coding sequence was isolated from plasmid pBT773 and inserted into the Nco I plus Kpn I digested KTI3 expression cassette described in Example 9.

EXAMPLE 12

Evaluating Compounds for Their Ability to Inhibit the Activity of Methionine Synthase

The plant methionine synthases described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 2, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant plant methionine synthases may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("$(His)_6$"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant plant methionine synthases, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the instant plant methionine synthase is expressed as fusion protein, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, the instant plant methionine synthase may be

expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a $(His)_6$ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β–mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the instant plant methionine synthase disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for methionine synthase are presented by Eichel et al. (1995) *Eur J Biochem 230*:1053-1058.